

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2004, 08:03:43 ; Search time 49.0945 Seconds  
(without alignments)  
247.473 Million cell updates/sec

Title: US-09-936-697-5  
Perfect score: 212  
Sequence: 1 PMRSISENSLVAMDFSGQKS.....ENPTEALSVAVEEGLAWRKK 43

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 842883

Minimum DB seq length: 0  
Maximum DB seq length: 85

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	% Query		DB	ID	Description
	Score	Match Length			
1	212	100.0	43	3 AAB18941	Aab18941 Peptide d
2	212	100.0	84	3 AAB18942	Aab18942 Peptide d
3	205	96.7	43	3 AAB18937	Aab18937 Peptide d
4	205	96.7	84	3 AAB18938	Aab18938 Peptide d
5	169	79.7	43	3 AAB18949	Aab18949 Peptide d
6	169	79.7	82	3 AAB18950	Aab18950 Peptide d
7	162	76.4	43	3 AAB18957	Aab18957 Peptide d
8	162	76.4	80	3 AAB18958	Aab18958 Peptide d
9	161	75.9	43	3 AAB18945	Aab18945 Peptide d

10	161	75.9	82	3	AAB18946	Aab18946	Peptide d
11	159	75.0	43	3	AAB18953	Aab18953	Peptide d
12	159	75.0	43	3	AAB18961	Aab18961	Peptide d
13	159	75.0	80	3	AAB18962	Aab18962	Peptide d
14	159	75.0	80	3	AAB18954	Aab18954	Peptide d
15	49	23.1	67	5	ABP08708	Abp08708	Human ORF
16	48	22.6	72	4	AAU30892	Aau30892	Novel hum
17	47.5	22.4	73	4	AAG76197	Aag76197	Human col
18	45.5	21.5	57	5	ABP02324	Abp02324	Human ORF
19	45	21.2	84	4	ABG59890	Abg59890	Human liv
20	45	21.2	84	5	ABG47266	Abg47266	Human pep
21	44.5	21.0	38	3	AAB38233	Aab38233	Human sec
22	44.5	21.0	72	3	AAG03340	Aag03340	Human sec
23	44.5	21.0	74	4	AAU87164	Aau87164	Novel cen
24	44.5	21.0	74	4	AAU87480	Aau87480	Novel cen
25	44.5	21.0	79	4	AAU41349	Aau41349	Propionib
26	44.5	21.0	79	6	ABM37868	Abm37868	Propionib
27	44	20.8	47	4	AAM94366	Aam94366	Human rep
28	44	20.8	62	4	ABG05569	Abg05569	Novel hum
29	44	20.8	80	2	AAW87732	Aaw87732	RSPaV str
30	44	20.8	82	4	ABB64548	Abb64548	Drosophil
31	43	20.3	14	4	AAM98073	Aam98073	Human pep
32	42	19.8	45	4	AAM87183	Aam87183	Human imm
33	42	19.8	63	4	AAM14547	Aam14547	Peptide #
34	42	19.8	63	4	ABB33505	Abb33505	Peptide #
35	42	19.8	63	4	AAM26965	Aam26965	Peptide #
36	42	19.8	63	4	ABB28328	Abb28328	Human pep
37	42	19.8	63	4	ABB18964	Abb18964	Protein #
38	42	19.8	63	4	AAM66679	Aam66679	Human bon
39	42	19.8	63	4	AAM54284	Aam54284	Human bra
40	42	19.8	63	4	ABG48351	Abg48351	Human liv
41	42	19.8	63	4	AAM02273	Aam02273	Peptide #
42	42	19.8	63	5	ABG36334	Abg36334	Human pep
43	42	19.8	64	4	AAM85610	Aam85610	Human imm
44	42	19.8	68	7	ADC95929	Adc95929	E. faeciu
45	42	19.8	70	6	AAE37094	Aae37094	Human DHA
46	42	19.8	70	6	ABU35221	Abu35221	Protein e
47	42	19.8	72	4	AAG93010	Aag93010	C glutami
48	42	19.8	76	6	ABP78540	Abp78540	N. gonorr
49	42	19.8	76	6	ABU37598	Abu37598	Protein e
50	42	19.8	81	2	AAY35749	Aay35749	Chlamydia
51	41.5	19.6	51	4	AAM96493	Aam96493	Human rep
52	41.5	19.6	71	4	AAU56590	Aau56590	Propionib
53	41.5	19.6	71	6	ABM53109	Abm53109	Propionib
54	41.5	19.6	83	4	ABB16786	Abb16786	Human ner
55	41	19.3	31	3	AAB11963	Aab11963	Human cat
56	41	19.3	52	4	AAU39844	Aau39844	Propionib
57	41	19.3	52	6	ABM36363	Abm36363	Propionib
58	41	19.3	61	7	ADC97267	Adc97267	E. faeciu
59	41	19.3	63	5	ABP00735	Abp00735	Human ORF
60	40.5	19.1	51	3	AAB16822	Aab16822	Bacteriop
61	40.5	19.1	66	3	AAB12066	Aab12066	SH3 domai
62	40.5	19.1	66	7	ADE84639	Ade84639	P85 SH3 b
63	40.5	19.1	74	3	AAG55105	Aag55105	Arabidops
64	40.5	19.1	79	6	ABU21800	Abu21800	Protein e
65	40.5	19.1	82	3	AAG03407	Aag03407	Human sec
66	40	18.9	36	4	AAM89235	Aam89235	Human imm

67	40	18.9	49	4	ABB42761	Abb42761	Peptide #
68	40	18.9	49	4	AAM36573	Aam36573	Peptide #
69	40	18.9	49	4	ABB26054	Abb26054	Protein #
70	40	18.9	49	4	AAM76466	Aam76466	Human bon
71	40	18.9	49	4	AAM63652	Aam63652	Human bra
72	40	18.9	49	4	ABG58167	Abg58167	Human liv
73	40	18.9	49	5	ABG45727	Abg45727	Human pep
74	40	18.9	57	3	AAG02206	Aag02206	Human sec
75	40	18.9	63	2	AAy02955	Aay02955	Fragment
76	40	18.9	63	7	ADA07790	Ada07790	Human sec
77	40	18.9	70	2	AAW98525	Aaw98525	H. pylori
78	40	18.9	71	4	AAM92537	Aam92537	Human dig
79	40	18.9	71	4	AAU45158	Aau45158	Propionib
80	40	18.9	71	4	AAU22574	Aau22574	Novel hum
81	40	18.9	71	6	ABM41677	Abm41677	Propionib
82	40	18.9	71	7	ADB32414	Adb32414	Human nov
83	40	18.9	72	3	AAG54855	Aag54855	Arabidops
84	40	18.9	74	3	AAG61701	Aag61701	Arabidops
85	39.5	18.6	54	2	AAy06664	Aay06664	Magnaport
86	39.5	18.6	61	5	ABB53884	Abb53884	Lactococc
87	39.5	18.6	68	4	AAM33238	Aam33238	Peptide #
88	39.5	18.6	68	4	AAM73013	Aam73013	Human bon
89	39.5	18.6	68	4	ABG54725	Abg54725	Human liv
90	39.5	18.6	68	5	ABG42853	Abg42853	Human pep
91	39.5	18.6	73	2	AAy08413	Aay08413	P. obesus
92	39.5	18.6	73	3	AAB36290	Aab36290	Israeli s
93	39.5	18.6	73	4	AAM39680	Aam39680	Human pol
94	39.5	18.6	78	4	AAM92935	Aam92935	Human dig
95	39.5	18.6	78	4	AAU20031	Aau20031	Human liv
96	39.5	18.6	78	5	ABP40892	Abp40892	Human liv
97	39.5	18.6	83	4	ABG28281	Abg28281	Novel hum
98	39	18.4	38	4	AAG99869	Aag99869	ERA bindi
99	39	18.4	40	4	AAG99833	Aag99833	ERA bindi
100	39	18.4	42	4	AAG99797	Aag99797	ERA bindi

# ALIGNMENTS

## RESULT 1

AAB18941

ID AAB18941 standard; peptide; 43 AA.

XX

AC AAB18941;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;

KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;

KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Homo sapiens.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.  
 XX  
 PF 14-MAR-2000; 2000WO-FR000613.  
 XX  
 PR 15-MAR-1999; 99FR-00003159.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX  
 DR WPI; 2000-587566/55.  
 XX  
 PT Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity.  
 XX  
 PS Claim 2; Page 25; 46pp; French.  
 XX  
 CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein. PIR  
 CC is the actual binding region but its effect is about 10 times greater in  
 CC presence of SH2 (which by itself is inactive). Agents that affect binding  
 CC between the peptides and the insulin receptor can stimulate or inhibit  
 CC tyrosine kinase activity of the receptor. The peptides are used for  
 CC screening molecules for ability to treat diseases in which insulin is  
 CC implicated. The peptides are used to identify agents that are potentially  
 CC useful for treating insulin-associated diseases, particularly diabetes  
 CC and obesity but also polycystic ovarian syndrome and syndrome X  
 XX  
 SQ Sequence 43 AA;

Query Match 100.0%; Score 212; DB 3; Length 43;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-24;  
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 ||||||||||||||||||||||||||||||||||||||||  
 Db 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43

# RESULT 2

AAB18942

ID AAB18942 standard; peptide; 84 AA.

XX

AC AAB18942;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Homo sapiens.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.  
 XX  
 PF 14-MAR-2000; 2000WO-FR000613.  
 XX  
 PR 15-MAR-1999; 99FR-00003159.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX  
 DR WPI; 2000-587566/55.  
 XX  
 PT Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity.  
 XX  
 PS Claim 2; Page 26; 46pp; French.  
 XX  
 CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein. PIR  
 CC is the actual binding region but its effect is about 10 times greater in  
 CC presence of SH2 (which by itself is inactive). Agents that affect binding  
 CC between the peptides and the insulin receptor can stimulate or inhibit  
 CC tyrosine kinase activity of the receptor. The peptides are used for  
 CC screening molecules for ability to treat diseases in which insulin is  
 CC implicated. The peptides are used to identify agents that are potentially  
 CC useful for treating insulin-associated diseases, particularly diabetes  
 CC and obesity but also polycystic ovarian syndrome and syndrome X  
 XX  
 SQ Sequence 84 AA;

Query Match 100.0%; Score 212; DB 3; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-24;  
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 ||||||||||||||||||||||||||||||||||||||||  
 Db 13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 55

RESULT 3

AAB18937

ID AAB18937 standard; peptide; 43 AA.

XX

AC AAB18937;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Rattus sp.

XX

PN WO200055634-A1.

XX



PD 21-SEP-2000.  
 XX  
 PF 14-MAR-2000; 2000WO-FR000613.  
 XX  
 PR 15-MAR-1999; 99FR-00003159.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX  
 DR WPI; 2000-587566/55.  
 XX  
 PT Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity.  
 XX  
 PS Claim 2; Page 23-24; 46pp; French.  
 XX  
 CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein. PIR  
 CC is the actual binding region but its effect is about 10 times greater in  
 CC presence of SH2 (which by itself is inactive). Agents that affect binding  
 CC between the peptides and the insulin receptor can stimulate or inhibit  
 CC tyrosine kinase activity of the receptor. The peptides are used for  
 CC screening molecules for ability to treat diseases in which insulin is  
 CC implicated. The peptides are used to identify agents that are potentially  
 CC useful for treating insulin-associated diseases, particularly diabetes  
 CC and obesity but also polycystic ovarian syndrome and syndrome X  
 XX  
 SQ Sequence 84 AA;

Query Match 96.7%; Score 205; DB 3; Length 84;  
 Best Local Similarity 93.0%; Pred. No. 3.6e-23;  
 Matches 40; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 ||||:|||||:||||:|||||  
 Db 13 PMRSVSENSLVAMDFSGQKTRVIDNPTEALSVAVEEGLAWRKK 55

# RESULT 5

AAB18949

ID AAB18949 standard; peptide; 43 AA.

XX

AC AAB18949;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;

KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;

KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Homo sapiens.

XX

PN WO200055634-A1.

XX







PD 21-SEP-2000.  
 XX  
 PF 14-MAR-2000; 2000WO-FR000613.  
 XX  
 PR 15-MAR-1999; 99FR-00003159.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX  
 DR WPI; 2000-587566/55.  
 XX  
 PT Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity.  
 XX  
 PS Claim 2; Page 34; 46pp; French.  
 XX  
 CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein. PIR  
 CC is the actual binding region but its effect is about 10 times greater in  
 CC presence of SH2 (which by itself is inactive). Agents that affect binding  
 CC between the peptides and the insulin receptor can stimulate or inhibit  
 CC tyrosine kinase activity of the receptor. The peptides are used for  
 CC screening molecules for ability to treat diseases in which insulin is  
 CC implicated. The peptides are used to identify agents that are potentially  
 CC useful for treating insulin-associated diseases, particularly diabetes  
 CC and obesity but also polycystic ovarian syndrome and syndrome X  
 XX  
 SQ Sequence 43 AA;

Query Match 76.4%; Score 162; DB 3; Length 43;  
 Best Local Similarity 74.4%; Pred. No. 5.4e-17;  
 Matches 32; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 |:|| |:|:||||||| ||||| |||||:|| |||||  
 Db 1 PLRSASDNTLVAMDFSGHAGRVIENTPREALSVALEEAQAWRKK 43

# RESULT 8

AAB18958

ID AAB18958 standard; peptide; 80 AA.

XX

AC AAB18958;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Homo sapiens.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.  
 XX  
 PF 14-MAR-2000; 2000WO-FR000613.  
 XX  
 PR 15-MAR-1999; 99FR-00003159.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX  
 DR WPI; 2000-587566/55.  
 XX  
 PT Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity.  
 XX  
 PS Claim 2; Page 34-35; 46pp; French.  
 XX  
 CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein. PIR  
 CC is the actual binding region but its effect is about 10 times greater in  
 CC presence of SH2 (which by itself is inactive). Agents that affect binding  
 CC between the peptides and the insulin receptor can stimulate or inhibit  
 CC tyrosine kinase activity of the receptor. The peptides are used for  
 CC screening molecules for ability to treat diseases in which insulin is  
 CC implicated. The peptides are used to identify agents that are potentially  
 CC useful for treating insulin-associated diseases, particularly diabetes  
 CC and obesity but also polycystic ovarian syndrome and syndrome X  
 XX  
 SQ Sequence 80 AA;

Query Match 76.4%; Score 162; DB 3; Length 80;  
 Best Local Similarity 74.4%; Pred. No. 1.3e-16;  
 Matches 32; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 PMRSISENSIVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 |:|| |:|:||||||| ||||| |||||:| |  
 Db 13 PLRSASDNTLVAMDFSGHAGRVIENTPREALSVALEEAQAWRKK 55

# RESULT 9

AAB18945

ID AAB18945 standard; peptide; 43 AA.

XX

AC AAB18945;

XX

DT 06-AUG-2003 (revised)

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Mus sp.

XX

PN WO200055634-A1.

XX  
 PD 21-SEP-2000.  
 XX  
 PF 14-MAR-2000; 2000WO-FR000613.  
 XX  
 PR 15-MAR-1999; 99FR-00003159.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX  
 DR WPI; 2000-587566/55.  
 XX  
 PT Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity.  
 XX  
 PS Claim 2; Page 27-28; 46pp; French.  
 XX  
 CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein. PIR  
 CC is the actual binding region but its effect is about 10 times greater in  
 CC presence of SH2 (which by itself is inactive). Agents that affect binding  
 CC between the peptides and the insulin receptor can stimulate or inhibit  
 CC tyrosine kinase activity of the receptor. The peptides are used for  
 CC screening molecules for ability to treat diseases in which insulin is  
 CC implicated. The peptides are used to identify agents that are potentially  
 CC useful for treating insulin-associated diseases, particularly diabetes  
 CC and obesity but also polycystic ovarian syndrome and syndrome X. (Updated  
 CC on 06-AUG-2003 to correct OS field.)  
 XX  
 SQ Sequence 43 AA;

Query Match 75.9%; Score 161; DB 3; Length 43;  
 Best Local Similarity 78.0%; Pred. No. 7.6e-17;  
 Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
 ||||:||||||| |||:| | | |:| | ||  
 Db 1 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGLHAWR 41

# RESULT 10

AAB18946

ID AAB18946 standard; peptide; 82 AA.

XX

AC AAB18946;

XX

DT 06-AUG-2003 (revised)

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Mus sp.

XX  
 PN WO200055634-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 14-MAR-2000; 2000WO-FR000613.  
 XX  
 PR 15-MAR-1999; 99FR-00003159.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX  
 DR WPI; 2000-587566/55.  
 XX  
 PT Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity.  
 XX  
 PS Claim 2; Page 28; 46pp; French.  
 XX  
 CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein. PIR  
 CC is the actual binding region but its effect is about 10 times greater in  
 CC presence of SH2 (which by itself is inactive). Agents that affect binding  
 CC between the peptides and the insulin receptor can stimulate or inhibit  
 CC tyrosine kinase activity of the receptor. The peptides are used for  
 CC screening molecules for ability to treat diseases in which insulin is  
 CC implicated. The peptides are used to identify agents that are potentially  
 CC useful for treating insulin-associated diseases, particularly diabetes  
 CC and obesity but also polycystic ovarian syndrome and syndrome X. (Updated  
 CC on 06-AUG-2003 to correct OS field.)  
 XX  
 SQ Sequence 82 AA;

Query Match 75.9%; Score 161; DB 3; Length 82;  
 Best Local Similarity 78.0%; Pred. No. 1.9e-16;  
 Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
 ||||:||||||| |||:| | | |:| | ||  
 Db 13 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWR 53

RESULT 11  
 AAB18953  
 ID AAB18953 standard; peptide; 43 AA.  
 XX  
 AC AAB18953;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
 XX  
 KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
 XX

OS Rattus sp.  
XX  
PN WO200055634-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 14-MAR-2000; 2000WO-FR000613.  
XX  
PR 15-MAR-1999; 99FR-00003159.  
XX  
PA (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;  
XX  
DR WPI; 2000-587566/55.  
XX  
PT Fragments of Grb family proteins to identify compounds are useful in  
PT treating insulin-associated diseases, particularly diabetes and obesity.  
XX  
PS Claim 2; Page 32; 46pp; French.  
XX  
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein. PIR  
CC is the actual binding region but its effect is about 10 times greater in  
CC presence of SH2 (which by itself is inactive). Agents that affect binding  
CC between the peptides and the insulin receptor can stimulate or inhibit  
CC tyrosine kinase activity of the receptor. The peptides are used for  
CC screening molecules for ability to treat diseases in which insulin is  
CC implicated. The peptides are used to identify agents that are potentially  
CC useful for treating insulin-associated diseases, particularly diabetes  
CC and obesity but also polycystic ovarian syndrome and syndrome X  
XX  
SQ Sequence 43 AA;

QY           1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
             |::||:|:|:|||||     ||:| | || | :| | ||||  
Db           1 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKK 43

ID AAB18961 standard; peptide; 43 AA.



KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Mus sp.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.

XX

PF 14-MAR-2000; 2000WO-FR000613.

XX

PR 15-MAR-1999; 99FR-00003159.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX

DR WPI; 2000-587566/55.

XX

PT Fragments of Grb family proteins to identify compounds are useful in  
PT treating insulin-associated diseases, particularly diabetes and obesity.

XX

PS Claim 2; Page 37; 46pp; French.

XX

CC Bl8937-64 represent the PIR (phosphorylated insulin receptor interacting  
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein. PIR  
CC is the actual binding region but its effect is about 10 times greater in  
CC presence of SH2 (which by itself is inactive). Agents that affect binding  
CC between the peptides and the insulin receptor can stimulate or inhibit  
CC tyrosine kinase activity of the receptor. The peptides are used for  
CC screening molecules for ability to treat diseases in which insulin is  
CC implicated. The peptides are used to identify agents that are potentially  
CC useful for treating insulin-associated diseases, particularly diabetes  
CC and obesity but also polycystic ovarian syndrome and syndrome X. (Updated  
CC on 06-AUG-2003 to correct OS field.)

XX

SQ Sequence 80 AA;

Query Match 75.0%; Score 159; DB 3; Length 80;

Best Local Similarity 69.8%; Pred. No. 3.6e-16;

Matches 30; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43

| : | : | : | : | | | | | | | | | | | | | : | | | | | | | |

Db 13 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKK 55

## RESULT 14

AAB18954

ID AAB18954 standard; peptide; 80 AA.

XX

AC    AAB18954;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX



KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS     Rattus sp.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.

XX

PF 14-MAR-2000; 2000WO-FR000613.

XX

PR 15-MAR-1999; 99FR-00003159.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX

DR WPI; 2000-587566/55.

XX

PT Fragments of Grb family proteins to identify compounds are useful in  
PT treating insulin-associated diseases, particularly diabetes and obesity.

XX

PS Claim 2; Page 32; 46pp; French.

XX

CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein. PIR  
CC is the actual binding region but its effect is about 10 times greater in  
CC presence of SH2 (which by itself is inactive). Agents that affect binding  
CC between the peptides and the insulin receptor can stimulate or inhibit  
CC tyrosine kinase activity of the receptor. The peptides are used for  
CC screening molecules for ability to treat diseases in which insulin is  
CC implicated. The peptides are used to identify agents that are potentially  
CC useful for treating insulin-associated diseases, particularly diabetes  
CC and obesity but also polycystic ovarian syndrome and syndrome X

XX

S0 Sequence 80 AA;

Query Match 75.0%; Score 159; DB 3; Length 80;

Best Local Similarity 69.8%; Pred. No. 3.6e-16;

Matches 30; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43

| : | : | : | : | |||||      ||||:|    |||||    :|    |||||

Db 13 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKK 55

RESULT 15

ABP08708

ID ABP08708 standard; protein; 67 AA.

XX

AC ABP08708;

XX

DT 24-JUN-2002 (first entry)

XX

DE Human ORFX protein sequence SEQ ID NO:17398.

XX

KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KW hypertension; hypothyroidism; cholesterol ester storage disease;  
 KW immune deficiency; immune disorder; infectious disease;  
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KW myasthenia gravis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200192523-A2.  
 XX  
 PD 06-DEC-2001.  
 XX  
 PF 29-MAY-2001; 2001WO-US010836.  
 XX  
 PR 30-MAY-2000; 2000US-0206132P.  
 PR 29-AUG-2000; 2000US-0228716P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shimkets RA, Leach MD;  
 XX  
 DR WPI; 2002-106308/14.  
 DR N-PSDB; ABN24460.  
 XX  
 PT Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and autoimmune disorders.  
 XX  
 PS Disclosure; SEQ ID NO 17398; 1037pp; English.  
 XX  
 CC The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
 CC treating or preventing a pathology associated with an ORFX-associated  
 CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
 CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage. N.B. The sequence data for this patent did not  
 CC form part of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX  
SQ Sequence 67 AA;

Query Match 23.1%; Score 49; DB 5; Length 67;  
Best Local Similarity 33.3%; Pred. No. 18;  
Matches 12; Conservative 3; Mismatches 15; Indels 6; Gaps 1;

Qy 11 VAMDFSGQKSRVIEN-----PTEALSVAVEEGLAW 40  
: ||| |||: | | |: | |  
Db 17 LGFSFSGPKSRVLSTSLHCPMPVEVLAEKEHGGFQW 52

RESULT 16  
AAU30892

ID AAU30892 standard; protein; 72 AA.

XX

AC AAU30892;

XX

DT 18-DEC-2001 (first entry)

XX

DE Novel human secreted protein #1383.

XX

KW Human; vaccination; gene therapy; nutritional supplement;

KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;

KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX

OS Homo sapiens.

XX

PN WO200179449-A2.

XX

PD 25-OCT-2001.

XX

PF 16-APR-2001; 2001WO-US008656.

XX

PR 18-APR-2000; 2000US-00552929.

PR 26-JAN-2001; 2001US-00770160.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Drmanac RT;

XX

DR WPI; 2001-611725/70.

XX

PT Nucleic acids encoding a range of human polypeptides, useful in genetic  
PT vaccination, testing and therapy.

XX

PS Claim 20; Page 366; 765pp; English.

XX

CC The invention relates to novel human secreted polypeptides. The  
CC polypeptides and antibodies to the polypeptides are useful for  
CC determining the presence of or predisposition to a disease associated  
CC with altered levels of polypeptide. The polypeptides are also useful for  
CC identifying agents (agonists and antagonists) that bind to them. Cells  
CC expressing the proteins are useful for identifying a therapeutic agent  
CC for use in treatment of a pathology related to aberrant expression or  
CC physiological interactions of the polypeptide. Vectors comprising the  
CC nucleic acids encoding the polypeptides and cells genetically engineered

CC to express them are also useful for producing the proteins. The proteins  
CC are useful in genetic vaccination, testing and therapy, and can be used  
CC as nutritional supplements. They may be used to increase stem cell  
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon  
CC and/or nerve tissue growth or regeneration; immune suppression and/or  
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.  
CC AAU29510-AAU33304 represent the amino acid sequences of novel human  
CC secreted proteins of the invention

XX

SQ Sequence 72 AA;

Query Match 22.6%; Score 48; DB 4; Length 72;  
Best Local Similarity 26.8%; Pred. No. 28;  
Matches 15; Conservative 13; Mismatches 14; Indels 14; Gaps 3;

QY 1 PMRSISENSLVAMDFSGQKSRV-----IENPTEALSVA-----VEEGLAWRKK 43  
|:| | : :: | :| : : ||: |:| :|||:|:|  
Db 16 PLSSXXLNKIPSLPSSWEKWXIPKNNCLSLNPSPP-SLAPSLDDIKEGLSWKKK 70

#### RESULT 17

AAG76197

ID AAG76197 standard; protein; 73 AA.

XX

AC AAG76197;

XX

DT 03-SEP-2001 (first entry)

XX

DE Human colon cancer antigen protein SEQ ID NO:6961.

XX

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma.

XX

OS Homo sapiens.

XX

PN WO200122920-A2.

XX

PD 05-APR-2001.

XX

PF 28-SEP-2000; 2000WO-US026524.

XX

PR 29-SEP-1999; 99US-0157137P.

PR 03-NOV-1999; 99US-0163280P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX

DR WPI; 2001-235357/24.

DR N-PSDB; AAH35602.

XX

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
PT useful for preventing, diagnosing and/or treating colorectal cancers.

XX

PS Claim 11; Page 8390; 9803pp; English.

XX

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon

CC cancer-associated nucleic acid molecules (N) and proteins (P), where the  
CC proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene therapy  
CC and vaccine production. N and P may be used in the prevention, diagnosis  
CC and treatment of diseases associated with inappropriate P expression. For  
CC example, N and P may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of P by expressing inactive proteins or to  
CC supplement the patients own production of P. Additionally, N may be used  
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids  
CC into a host cell and culturing the cell to express the proteins. N and P  
CC can be used in the prevention, diagnosis and treatment of colorectal  
CC carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent  
CC sequences used in the exemplification of the present invention. N.B.  
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at  
CC time of publication, meaning no sequences are present for SEQ ID NO:1027  
CC to 1052, 7921 and 7922

XX

SQ Sequence 73 AA;

Query Match 22.4%; Score 47.5; DB 4; Length 73;  
Best Local Similarity 30.4%; Pred. No. 35;  
Matches 14; Conservative 8; Mismatches 17; Indels 7; Gaps 2;

Qy 4 SISENSLVAMDFSGQKSRVIE-----NPTEAL--SVAVEEGLAWRK 42  
:|||| :|: : : :| | |||| ||:| :  
Db 11 TISENLFATTGYPGKMASQFIHHLGHPQPILMGSAVAVGSGLSWHR 56

#### RESULT 18

ABP02324

ID ABP02324 standard; protein; 57 AA.

XX

AC ABP02324;

XX

DT 24-JUN-2002 (first entry)

XX

DE Human ORFX protein sequence SEQ ID NO:4630.

XX

KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis.

XX

OS Homo sapiens.

XX

PN WO200192523-A2.

XX

PD 06-DEC-2001.

XX

PF 29-MAY-2001; 2001WO-US010836.

XX

PR 30-MAY-2000; 2000US-0206132P.

PR 29-AUG-2000; 2000US-0228716P.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Shimkets RA, Leach MD;

XX

DR WPI; 2002-106308/14.

DR N-PSDB; ABN18076.

XX

PT Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders.

XX

PS Disclosure; SEQ ID NO 4630; 1037pp; English.

XX

CC The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage. N.B. The sequence data for this patent did not  
CC form part of the printed specification, but was obtained in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 57 AA;

Query Match 21.5%; Score 45.5; DB 5; Length 57;

Best Local Similarity 36.4%; Pred. No. 50;

Matches 12; Conservative 7; Mismatches 5; Indels 9; Gaps 2;

Qy 15 FSGQKSRVIENPT-----EALSVAVEEGLAWR 41

:||| |:| | |:|::: |:|

Db 17 WSGQ---VLENAVRWGLRREPLNVSLQNGKSWR 46

RESULT 19

ABG59890

ID ABG59890 standard; peptide; 84 AA.

XX

AC ABG59890;

XX

DT 25-FEB-2003 (first entry)  
 XX  
 DE Human liver peptide, SEQ ID No 38538.  
 XX  
 KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
 KW hypercholesterolaemia; coronary heart disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157273-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000664.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488898/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human adult liver.  
 XX  
 PS Claim 27; SEQ ID NO 38538; 658pp; English.  
 XX  
 CC The invention relates to a single exon nucleic acid probe (SENP) (I) for  
 CC measuring human gene expression in a sample derived from human adult  
 CC liver, comprising one of 13109 defined nucleotide sequences given in the  
 CC specification (or complements/ fragments). The probe hybridises at high  
 CC stringency to a nucleic acid molecule expressed in the human adult liver.  
 CC (I) may be used for predicting, measuring and displaying gene expression  
 CC in samples derived from human adult liver. The genes identified may be  
 CC involved in genetic liver diseases such as cirrhosis,  
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
 CC associated with coronary heart disease. ABG47348-ABG59930 represent human  
 CC liver single exon encoded peptides of the invention. Note: The sequence  
 CC information for this patent does not appear in the printed specification  
 CC but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 84 AA;

Query Match 21.2%; Score 45; DB 4; Length 84;  
 Best Local Similarity 26.4%; Pred. No. 1e+02;  
 Matches 14; Conservative 6; Mismatches 7; Indels 26; Gaps 2;

Qy 17 GQKSRVIENP-----TEALSVAV-----EEGLAWRKK 43  
 |||::: | | ||: |||:::

## RESULT 20

ABG47266

ID ABG47266 standard; peptide; 84 AA.

XX

AC ABG47266;

XX

DT 19-AUG-2002 (first entry)

XX

DE Human peptide encoded by genome-derived single exon probe SEQ ID 36931.

XX

KW Human; single exon probe; asthma; lung cancer; COPD; ILD;

KW chronic obstructive pulmonary disease; interstitial lung disease;

KW familial idiopathic pulmonary fibrosis; neurofibromatosis;

KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;

KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;

KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

KW primary ciliary dyskinesia; pulmonary hypertension;

KW hyaline membrane disease.

XX

OS Homo sapiens.

XX

PN WO200186003-A2.

XX

PD 15-NOV-2001.

XX

PF 30-JAN-2001; 2001WO-US000665.

XX

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2002-114183/15.

XX

PT Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples.

XX

PS Claim 27; SEQ ID NO 36931; 634pp; English.

XX

CC The invention relates to a spatially-addressable set of single exon

CC nucleic acid probes for measuring gene expression in a sample derived

CC from human lung comprising single exon nucleic acid probes having one of

CC 12614 nucleic acid sequences mentioned in the specification, or their

CC complements or the 12387 open reading frames derived from the 12614

CC probes. Also included are a microarray comprising the novel set of probes

CC ; the novel set of probes which hybridise at high stringency to a nucleic



CC acid expressed in the human lung; measuring gene expression in a sample  
 CC derived from human lung, comprising (a) contacting the array with a  
 CC collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of the  
 CC array; identifying exons in a eukaryotic genome, comprising (a)  
 CC algorithmically predicting at least one exon from genomic sequences of  
 CC the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene expression  
 CC analysis, and for identifying exons in a gene, particularly using human  
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
 CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
 CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
 CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
 CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
 CC Karagen syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
 CC present sequence is a peptide/protein encoded by a single exon probe of  
 CC the invention. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 84 AA;

Query Match 21.2%; Score 45; DB 5; Length 84;  
 Best Local Similarity 26.4%; Pred. No. 1e+02;  
 Matches 14; Conservative 6; Mismatches 7; Indels 26; Gaps 2;

QY 17 GQKSRVIENP-----TEALSVAV-----EEGLAWRKK 43  
 |||:|:: | | ||: || | ||::  
 Db 11 GQKARLLSRPLRGVSGKHCLTFFYHMYGGGTGLLSVYLKKEEDSEESLLWRRR 63

# RESULT 21

AAB38233

ID AAB38233 standard; protein; 38 AA.

XX

AC AAB38233;

XX

DT 30-JAN-2001 (first entry)

XX

DE Human secreted protein sequence encoded by gene 31 SEQ ID NO:89.

XX

KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;

KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;

KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;

KW fungicide; ophthalmological; gene therapy; autoimmune disease; infection;

KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;  
 KW cerebrovascular disorder; nervous system disorder; ocular disorder;  
 KW wound healing; skin aging; food additive; preservative.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200058469-A1.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 23-MAR-2000; 2000WO-US007579.  
 XX  
 PR 26-MAR-1999; 99US-0126509P.  
 PR 07-JAN-2000; 2000US-0174853P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM, Komatsoulis G;  
 XX  
 DR WPI; 2000-594642/56.  
 DR N-PSDB; AAC69485.  
 XX  
 PT Isolated nucleic acid molecule encoding a human secreted protein is used  
 PT in preventing, treating or ameliorating a medical condition.  
 XX  
 PS Claim 11; Page 370; 416pp; English.  
 XX  
 CC The polynucleotide sequences given in AAC69455 to AAC69502 encode the  
 CC human secreted proteins given in AAB38203 to AAB38250. AAB38251 to  
 CC AAB38320 represent human secreted polypeptide sequences and proteins  
 CC homologous to them, which are given in the exemplification of the present  
 CC invention. Human secreted proteins have activities based on the tissues  
 CC and cells the genes are expressed in. Example of activities include:  
 CC immunosuppressive; antiarthritic; antirheumatic; antiproliferative;  
 CC cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;  
 CC neuroprotective; antibacterial; virucide; fungicide; and  
 CC ophthalmological. The polynucleotides and polypeptides can be used to  
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
 CC in diagnosing a pathological condition or susceptibility to a  
 CC pathological condition. Disorders which are diagnosed or treated include  
 CC autoimmune diseases, hyperproliferative disorders, cardiovascular  
 CC disorders, cerebrovascular disorders, angiogenesis, nervous system  
 CC disorders, infections caused by bacteria, viruses and fungi and ocular  
 CC disorders. The polypeptides can also be used to aid wound healing and  
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
 CC maintain organs before transplantation, for supporting cell culture of  
 CC primary tissues, to regenerate tissues and in chemotaxis. The  
 CC polypeptides can also be used as a food additive or preservative to  
 CC increase or decrease storage capabilities. AAC69446 to AAC69454 and  
 CC AAB38202 represent sequences used in the exemplification of the present  
 CC invention  
 XX  
 SQ Sequence 38 AA;

Query Match 21.0%; Score 44.5; DB 3; Length 38;  
 Best Local Similarity 40.0%; Pred. No. 40;

Matches 10; Conservative 7; Mismatches 5; Indels 3; Gaps 1;

Qy 16 SGQKSRVIENPTEALSVAVEEGLAW 40  
:|: | |::: |: :| | |||  
Db 15 AGELSVVLQDSTDCMS---ELGLAW 36

RESULT 22

AAG03340

ID AAG03340 standard; protein; 72 AA.

XX

AC AAG03340;

XX

DT 06-OCT-2000 (first entry)

XX

DE Human secreted protein, SEQ ID NO: 7421.

XX

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping.

XX

OS Homo sapiens.

XX

PN EP1033401-A2.

XX

PD 06-SEP-2000.

XX

PF 21-FEB-2000; 2000EP-00200610.

XX

PR 26-FEB-1999; 99US-0122487P.

XX

PA (GEST ) GENSET.

XX

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX

DR WPI; 2000-500381/45.

DR N-PSDB; AAC03346.

XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX

PS Claim 13; SEQ ID NO 7421; 7lpp + Sequence Listing; English.

XX

CC The present sequence is a polypeptide encoded by one of a large number of  
CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were  
CC prepared from total human RNAs or polyA+ RNAs derived from 30 different  
CC tissues. EST sequences usually correspond mainly to the 3' untranslated  
CC region (UTR) of the mRNA because they are often obtained from oligo-dT  
CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA  
CC sequences derived from the 5' ends of mRNAs and even in those cases where  
CC longer cDNA sequences have been obtained, the full 5' UTR is rarely  
CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can  
CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs  
CC are also used in diagnostic, forensic, gene therapy and chromosome  
CC mapping procedures. They are used to obtain upstream regulatory sequences  
CC and to design expression and secretion vectors

XX

SQ Sequence 72 AA;

Query Match 21.0%; Score 44.5; DB 3; Length 72;  
Best Local Similarity 50.0%; Pred. No. 97;  
Matches 10; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

Qy 21 RVIENTEALSVAVEEGLAW 40  
|| :|||: ||| : |: |  
Db 38 RVCTHPTESCSVA-QAGVQW 56

RESULT 23

AAU87164

ID AAU87164 standard; protein; 74 AA.

XX

AC AAU87164;

XX

DT 05-JUN-2002 (first entry)

XX

DE Novel central nervous system protein #74.

XX

KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;

KW hyperproliferative disorder; neoplasm; cardiovascular disorder;

KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;

KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;

KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;

KW adenocarcinoma; reproductive system disorder; testicular feminisation;

KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;

KW respiratory disorder; renal disorder; kidney failure; blood disorder;

KW myocardial infarction; wound healing; cell proliferation; skin aging;

KW food additive; food preservative; gene therapy.

XX

OS Homo sapiens.

XX

PN WO200155318-A2.

XX

PD 02-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US001332.

XX

PR 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.

PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.

PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-581633/65.  
 DR N-PSDB; ABK43494.  
 XX  
 PT New isolated nucleic acid encoding a protein for diagnosing, preventing,  
 PT treating or ameliorating medical conditions and used as food additives or  
 PT preservatives.  
 XX  
 PS Claim 9; SEQ ID NO 682; 837pp; English.  
 XX  
 CC The invention describes an isolated nucleic acid molecule (I) encoding a  
 CC novel central nervous system protein. (I) and polypeptides (III) encoded  
 CC by (I), are used to treat a medical conditions and in diagnosis of a  
 CC pathological condition. Disorders which are diagnosed or treated include  
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative  
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders  
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,  
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and  
 CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses  
 CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders  
 CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,  
 CC adenocarcinomas and irritable bowel syndrome, reproductive system  
 CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes  
 CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.  
 CC leukaemia, disorders involving neovascularisation e.g. malignancies,  
 CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.  
 CC acute kidney failure and blood related disorders e.g. myocardial  
 CC infarction. The polypeptides can also be used to aid wound healing and  
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
 CC maintain organs before transplantation, for supporting cell culture of  
 CC primary tissues, to regenerate tissues and in chemotaxis. The  
 CC polypeptides can also be used as a food additive or preservative to  
 CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 21.0%; Score 44.5; DB 4; Length 74;  
 Best Local Similarity 23.6%; Pred. No. 1e+02;  
 Matches 13; Conservative 11; Mismatches 16; Indels 15; Gaps 2;

Qy 4 SISENSLVAMDFSGQKSRVIEN--PTEALSVA-----VEEGLAWRKK 43  
 ||:|| |: |: |:: :: | |: : : ||| |  
 Db 20 SITENGLIPKDYRSLKTQYLQSYGPEHLLTFSNLRXAGLLTXQAPGDNXTAWRVK 74

RESULT 24  
 AAU87480  
 ID AAU87480 standard; protein; 74 AA.  
 XX  
 AC AAU87480;

XX  
 DT 05-JUN-2002 (first entry)  
 XX  
 DE Novel central nervous system protein #390.  
 XX  
 KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
 KW hyperproliferative disorder; neoplasm; cardiovascular disorder;  
 KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;  
 KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
 KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
 KW adenocarcinoma; reproductive system disorder; testicular feminisation;  
 KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;  
 KW respiratory disorder; renal disorder; kidney failure; blood disorder;  
 KW myocardial infarction; wound healing; cell proliferation; skin aging;  
 KW food additive; food preservative; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200155318-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US001332.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226868P.



PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.

PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Barash SC, Ruben SM;

XX

DR WPI; 2001-581633/65.

DR N-PSDB; ABK43810.

XX

PT New isolated nucleic acid encoding a protein for diagnosing, preventing,  
PT treating or ameliorating medical conditions and used as food additives or  
PT preservatives.

XX  
PS Claim 9; SEQ ID NO 998; 837pp; English.  
XX  
CC The invention describes an isolated nucleic acid molecule (I) encoding a  
CC novel central nervous system protein. (I) and polypeptides (III) encoded  
CC by (I), are used to treat a medical conditions and in diagnosis of a  
CC pathological condition. Disorders which are diagnosed or treated include  
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative  
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders  
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,  
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and  
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses  
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders  
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,  
CC adenocarcinomas and irritable bowel syndrome, reproductive system  
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes  
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.  
CC leukaemia, disorders involving neovascularisation e.g. malignancies,  
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.  
CC acute kidney failure and blood related disorders e.g. myocardial  
CC infarction. The polypeptides can also be used to aid wound healing and  
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
CC maintain organs before transplantation, for supporting cell culture of  
CC primary tissues, to regenerate tissues and in chemotaxis. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 21.0%; Score 44.5; DB 4; Length 74;  
Best Local Similarity 23.6%; Pred. No. 1e+02;  
Matches 13; Conservative 11; Mismatches 16; Indels 15; Gaps 2;

Qy 4 SISENSLVAMDFSGQKSRVIEN--PTEALSVA-----VEEGLAWRKK 43  
||:| | : | : : : | : : : | : :  
Db 20 SITENGLIPKDYRSLKTQYLQSYGPEHLLTFSNLRXAGLLTXQAPGDNXTAWRVK 74

# RESULT 25

AAU41349

ID AAU41349 standard; protein; 79 AA.

XX

AC AAU41349;

XX

DT 13-FEB-2002 (first entry)

XX

DE Propionibacterium acnes immunogenic protein #2245.

XX

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;

KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;

KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;

KW dermatological; osteopathic; neuroprotectant.

XX

OS Propionibacterium acnes.

XX

PN WO200181581-A2.

XX

PD 01-NOV-2001.

XX

PF 20-APR-2001; 2001WO-US012865.

XX

PR 21-APR-2000; 2000US-0199047P.

PR 02-JUN-2000; 2000US-0208841P.

PR 07-JUL-2000; 2000US-0216747P.

XX

PA (CORI-) CORIXA CORP.

XX

PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX

DR WPI; 2001-616774/71.

DR N-PSDB; AAS59515.

XX

PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris.

XX

PS Example 1; SEQ ID NO 2544; 1069pp; English.

XX

CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 79 AA;

Query Match 21.0%; Score 44.5; DB 4; Length 79;

Best Local Similarity 27.5%; Pred. No. 1.1e+02;

Matches 11; Conservative 10; Mismatches 16; Indels 3; Gaps 1;

Qy 4 SISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43

|: : : |: | :|| || :|: | |: |::

Db 28 SLVNSPVTALSREGPSNRV---PTRSLACATRHGVCSRER 64

Search completed: July 8, 2004, 08:19:13

Job time : 54.0945 secs

OM protein - protein search, using sw model

Run on: July 8, 2004, 08:16:33 ; Search time 10.8346 Seconds  
 (without alignments)  
 204.891 Million cell updates/sec

Title: US-09-936-697-5  
 Perfect score: 212  
 Sequence: 1 PMRSISENSLVAMDFSGQKS.....ENPTEALSVAVEEGLAWRKK 43

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 257387

Minimum DB seq length: 0  
 Maximum DB seq length: 85

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 100 summaries

Database : Issued\_Patents\_AA:\*  
 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
 5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

		%					
Result		Query					Description
No.	Score	Match	Length	DB	ID		
1	45.5	21.5	73	4	US-09-331-930A-22		Sequence 22, Appl
2	44.5	21.0	62	4	US-09-252-991A-32126		Sequence 32126, A
3	44	20.8	47	3	US-08-776-059-18		Sequence 18, Appl
4	44	20.8	80	3	US-09-081-320-20		Sequence 20, Appl
5	44	20.8	80	4	US-09-574-141A-20		Sequence 20, Appl
6	44	20.8	80	4	US-09-707-780-20		Sequence 20, Appl
7	43.5	20.5	55	4	US-08-630-915A-111		Sequence 111, App
8	43	20.3	81	4	US-09-134-000C-5090		Sequence 5090, Ap
9	42	19.8	47	3	US-08-776-059-16		Sequence 16, Appl
10	42	19.8	61	4	US-09-134-000C-5679		Sequence 5679, Ap
11	42	19.8	68	4	US-09-107-532A-5556		Sequence 5556, Ap

12	42	19.8	78	4	US-09-540-236-2395	Sequence 2395, Ap
13	42	19.8	81	4	US-09-198-452A-1167	Sequence 1167, Ap
14	41	19.3	40	1	US-07-641-971B-5	Sequence 5, Appli
15	41	19.3	40	1	US-07-781-248A-5	Sequence 5, Appli
16	41	19.3	61	4	US-09-107-532A-6894	Sequence 6894, Ap
17	41	19.3	79	4	US-09-621-976-7338	Sequence 7338, Ap
18	40.5	19.1	51	4	US-09-023-905A-17	Sequence 17, Appl
19	40.5	19.1	66	2	US-08-459-568-52	Sequence 52, Appl
20	40.5	19.1	66	2	US-08-399-411-52	Sequence 52, Appl
21	40.5	19.1	66	3	US-08-516-859A-52	Sequence 52, Appl
22	40.5	19.1	66	4	US-09-586-472-52	Sequence 52, Appl
23	40.5	19.1	66	4	US-09-528-706-52	Sequence 52, Appl
24	40.5	19.1	71	4	US-09-621-976-5251	Sequence 5251, Ap
25	40.5	19.1	79	4	US-09-006-428A-14	Sequence 14, Appl
26	40	18.9	63	4	US-09-227-357-611	Sequence 611, App
27	39.5	18.6	24	6	5240706-21	Patent No. 5240706
28	39.5	18.6	73	4	US-09-331-930A-2	Sequence 2, Appli
29	39.5	18.6	73	4	US-09-331-930A-19	Sequence 19, Appl
30	39.5	18.6	73	4	US-09-331-930A-20	Sequence 20, Appl
31	39.5	18.6	73	4	US-09-331-930A-21	Sequence 21, Appl
32	39.5	18.6	73	4	US-09-331-930A-24	Sequence 24, Appl
33	39.5	18.6	73	4	US-09-331-930A-25	Sequence 25, Appl
34	39	18.4	45	4	US-09-732-210-1554	Sequence 1554, Ap
35	39	18.4	62	4	US-09-543-681A-4511	Sequence 4511, Ap
36	39	18.4	63	4	US-09-489-039A-7967	Sequence 7967, Ap
37	39	18.4	68	3	US-09-100-804-30	Sequence 30, Appl
38	39	18.4	72	4	US-09-543-681A-5442	Sequence 5442, Ap
39	39	18.4	80	3	US-09-081-320-9	Sequence 9, Appli
40	39	18.4	80	4	US-09-574-141A-9	Sequence 9, Appli
41	39	18.4	80	4	US-09-574-141A-78	Sequence 78, Appl
42	39	18.4	80	4	US-09-707-780-9	Sequence 9, Appli
43	39	18.4	81	4	US-09-252-991A-18621	Sequence 18621, A
44	38.5	18.2	30	3	US-08-847-065-5	Sequence 5, Appli
45	38.5	18.2	43	4	US-09-209-676-66	Sequence 66, Appl
46	38.5	18.2	59	1	US-08-524-677-16	Sequence 16, Appl
47	38.5	18.2	81	4	US-09-489-039A-11502	Sequence 11502, A
48	38.5	18.2	85	4	US-09-621-976-4396	Sequence 4396, Ap
49	38	17.9	62	4	US-08-936-165A-301	Sequence 301, App
50	38	17.9	64	4	US-09-107-532A-6576	Sequence 6576, Ap
51	38	17.9	70	1	US-08-167-035-26	Sequence 26, Appl
52	38	17.9	70	1	US-08-208-887A-26	Sequence 26, Appl
53	38	17.9	70	2	US-08-539-005-26	Sequence 26, Appl
54	38	17.9	70	4	US-09-280-598-28	Sequence 28, Appl
55	38	17.9	80	3	US-09-081-320-31	Sequence 31, Appl
56	38	17.9	80	4	US-09-574-141A-31	Sequence 31, Appl
57	38	17.9	80	4	US-09-707-780-31	Sequence 31, Appl
58	38	17.9	84	6	5171684-3	Patent No. 5171684
59	37.5	17.7	49	4	US-09-079-030-16	Sequence 16, Appl
60	37.5	17.7	66	4	US-09-331-930A-23	Sequence 23, Appl
61	37.5	17.7	75	4	US-09-134-000C-3960	Sequence 3960, Ap
62	37	17.5	13	4	US-09-484-114B-17	Sequence 17, Appl
63	37	17.5	13	4	US-09-981-194-17	Sequence 17, Appl
64	37	17.5	46	4	US-08-492-723-13	Sequence 13, Appl
65	37	17.5	55	4	US-09-205-258-1215	Sequence 1215, Ap
66	37	17.5	57	3	US-09-177-249-294	Sequence 294, App
67	37	17.5	73	2	US-08-891-837B-2	Sequence 2, Appli
68	37	17.5	81	4	US-09-489-039A-13399	Sequence 13399, A

69	36.5	17.2	43	4	US-09-131-750-25	Sequence 25, Appl
70	36.5	17.2	61	4	US-08-630-915A-208	Sequence 208, App
71	36	17.0	35	1	US-08-038-343A-5	Sequence 5, Appli
72	36	17.0	36	3	US-09-045-764A-12	Sequence 12, Appl
73	36	17.0	47	3	US-08-776-059-17	Sequence 17, Appl
74	36	17.0	58	1	US-08-446-646-13	Sequence 13, Appl
75	36	17.0	64	4	US-09-543-681A-4984	Sequence 4984, Ap
76	36	17.0	66	1	US-08-203-806B-11	Sequence 11, Appl
77	36	17.0	66	2	US-08-414-657D-54	Sequence 54, Appl
78	36	17.0	66	2	US-08-414-657D-55	Sequence 55, Appl
79	36	17.0	66	4	US-09-017-754A-11	Sequence 11, Appl
80	36	17.0	76	1	US-08-667-679-2	Sequence 2, Appli
81	36	17.0	76	2	US-08-892-150-2	Sequence 2, Appli
82	36	17.0	78	4	US-09-134-000C-5769	Sequence 5769, Ap
83	35.5	16.7	56	3	US-08-973-462-21	Sequence 21, Appl
84	35.5	16.7	56	3	US-08-973-462-24	Sequence 24, Appl
85	35.5	16.7	71	4	US-09-543-681A-5866	Sequence 5866, Ap
86	35.5	16.7	78	4	US-09-254-352B-23	Sequence 23, Appl
87	35	16.5	59	4	US-09-621-976-5299	Sequence 5299, Ap
88	35	16.5	67	4	US-09-621-976-7240	Sequence 7240, Ap
89	35	16.5	80	4	US-09-134-000C-6418	Sequence 6418, Ap
90	35	16.5	81	4	US-09-489-039A-14245	Sequence 14245, A
91	35	16.5	84	4	US-09-424-978B-26	Sequence 26, Appl
92	35	16.5	85	4	US-09-075-460-2	Sequence 2, Appli
93	34.5	16.3	58	1	US-08-222-616-30	Sequence 30, Appl
94	34.5	16.3	58	4	US-08-446-648-30	Sequence 30, Appl
95	34.5	16.3	58	4	US-09-982-610-30	Sequence 30, Appl
96	34.5	16.3	58	5	PCT-US95-04228-30	Sequence 30, Appl
97	34.5	16.3	82	4	US-09-252-991A-16574	Sequence 16574, A
98	34.5	16.3	82	4	US-09-252-991A-24955	Sequence 24955, A
99	34.5	16.3	83	3	US-09-100-804-24	Sequence 24, Appl
100	34.5	16.3	84	4	US-09-107-532A-5121	Sequence 5121, Ap

#### ALIGNMENTS

##### RESULT 1

US-09-331-930A-22

```
; Sequence 22, Application US/09331930A
; Patent No. 6436670
; GENERAL INFORMATION:
; APPLICANT: ZIMMET, PAUL Z.
; APPLICANT: COLLIER, GREGORY
; TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR
; FILE REFERENCE: 22975-20007.00
; CURRENT APPLICATION NUMBER: US/09/331,930A
; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: PCT/AU98/00902
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: AU PP0117/97
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: AU PP0323/97
; PRIOR FILING DATE: 1997-11-11
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
```

; LENGTH: 73  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-09-331-930A-22

Query Match 21.5%; Score 45.5; DB 4; Length 73;  
Best Local Similarity 29.4%; Pred. No. 13;  
Matches 10; Conservative 7; Mismatches 12; Indels 5; Gaps 1;

Qy 14 DFSGQKSRVIENPTEALS-----VAVEEGLAWRK 42  
| |:| |: ||:: : :| :| ||  
Db 8 DRLGKKVRIKCNPSDTIGDLKKLIAAQTGTRWEK 41

RESULT 2

US-09-252-991A-32126  
; Sequence 32126, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 32126  
; LENGTH: 62  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32126

Query Match 21.0%; Score 44.5; DB 4; Length 62;  
Best Local Similarity 40.0%; Pred. No. 15;  
Matches 10; Conservative 4; Mismatches 8; Indels 3; Gaps 1;

Qy 20 SRVIENPTE---ALSVAVEEGLAWR 41  
|| |:|| |:| :|| |:  
Db 32 SRTPEHPTSCACAISYKIFEGFCWK 56

RESULT 3

US-08-776-059-18  
; Sequence 18, Application US/08776059B  
; Patent No. 6271368  
; GENERAL INFORMATION:  
; APPLICANT: LENTZEN, Hans  
; APPLICANT: ECK, Jurgen  
; APPLICANT: BAUR, Axel  
; APPLICANT: ZINKE, Holger  
; TITLE OF INVENTION: RECOMBINANT MISTLETOE LECTIN (RML)  
; FILE REFERENCE: 674503-2003



; CURRENT APPLICATION NUMBER: US/08/776,059B  
; CURRENT FILING DATE: 1999-06-19  
; EARLIER APPLICATION NUMBER: PCT/EP96/02273  
; EARLIER FILING DATE: 1996-06-25  
; EARLIER APPLICATION NUMBER: 95109949.8  
; EARLIER FILING DATE: 1995-06-26  
; NUMBER OF SEQ ID NOS: 56  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 47  
; TYPE: PRT  
; ORGANISM: Saponaria officinalis  
US-08-776-059-18

Query Match 20.8%; Score 44; DB 3; Length 47;  
Best Local Similarity 34.8%; Pred. No. 12;  
Matches 8; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 13 MDFSGQKSRVIENPTEALSVAVE 35  
|| :|:|:| | :|:  
Db 6 MDAVNKKARVVKNEARFLIAIQ 28

RESULT 4

US-09-081-320-20

; Sequence 20, Application US/09081320  
; Patent No. 6093544  
; GENERAL INFORMATION:  
; APPLICANT: Gonsalves, Dennis  
; APPLICANT: Meng, Baozhong  
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS  
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: Clinton Square, P.O. Box 1051  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/081,320  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/047,147  
; FILING DATE: 20-MAY-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/069,902  
; FILING DATE: 17-DEC-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.

; REGISTRATION NUMBER: 30,727  
 ; REFERENCE/DOCKET NUMBER: 19603/1722  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (716) 263-1304  
 ; TELEFAX: (716) 263-1600  
 ; INFORMATION FOR SEQ ID NO: 20:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 80 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-09-081-320-20

Query Match 20.8%; Score 44; DB 3; Length 80;  
 Best Local Similarity 50.0%; Pred. No. 25;  
 Matches 13; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

Qy 19 KSRVIEN--PTEALSVAVEEGLAWRK 42  
 :| |||| |:|: |:| |  
 Db 40 ESIVIENCGPSEALAATVKEVLGGLK 65

# RESULT 5

US-09-574-141A-20  
 ; Sequence 20, Application US/09574141A  
 ; Patent No. 6395490  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gonsalves, Dennis  
 ; APPLICANT: Meng, Baozhong  
 ; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS  
 ; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES  
 ; FILE REFERENCE: 07678/035005  
 ; CURRENT APPLICATION NUMBER: US/09/574,141A  
 ; CURRENT FILING DATE: 2000-05-18  
 ; PRIOR APPLICATION NUMBER: 60/047,147  
 ; PRIOR FILING DATE: 1997-05-20  
 ; PRIOR APPLICATION NUMBER: 60/069,902  
 ; PRIOR FILING DATE: 1997-12-17  
 ; PRIOR APPLICATION NUMBER: 09/081,320  
 ; PRIOR FILING DATE: 1998-05-19  
 ; NUMBER OF SEQ ID NOS: 97  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 20  
 ; LENGTH: 80  
 ; TYPE: PRT  
 ; ORGANISM: Rupestris stem pitting associated virus  
 US-09-574-141A-20

Query Match 20.8%; Score 44; DB 4; Length 80;  
 Best Local Similarity 50.0%; Pred. No. 25;  
 Matches 13; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

Qy 19 KSRVIEN--PTEALSVAVEEGLAWRK 42  
 :| |||| |:|: |:| |  
 Db 40 ESIVIENCGPSEALAATVKEVLGGLK 65

RESULT 6

US-09-707-780-20

; Sequence 20, Application US/09707780  
 ; Patent No. 6399308  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gonsalves, Dennis  
 ; APPLICANT: Meng, Baozhong  
 ; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS  
 ; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES  
 ; FILE REFERENCE: 07678/035006  
 ; CURRENT APPLICATION NUMBER: US/09/707,780  
 ; CURRENT FILING DATE: 2000-11-07  
 ; PRIOR APPLICATION NUMBER: 09/081,320  
 ; PRIOR FILING DATE: 1998-05-19  
 ; PRIOR APPLICATION NUMBER: 60/047,147  
 ; PRIOR FILING DATE: 1997-05-20  
 ; PRIOR APPLICATION NUMBER: 60/069,902  
 ; PRIOR FILING DATE: 1997-12-17  
 ; NUMBER OF SEQ ID NOS: 54  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 20  
 ; LENGTH: 80  
 ; TYPE: PRT  
 ; ORGANISM: Rupestris stem pitting associated virus  
 US-09-707-780-20

Query Match 20.8%; Score 44; DB 4; Length 80;  
 Best Local Similarity 50.0%; Pred. No. 25;  
 Matches 13; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

Qy 19 KSRVIEN--PTEALSVAVEEGLAWRK 42  
 :| |||| |:|||: |:| | |  
 Db 40 ESIVIENC GPSEALAATVKEVLGGLK 65

RESULT 7

US-08-630-915A-111

; Sequence 111, Application US/08630915A  
 ; Patent No. 6309820  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SPARKS, Andrew B.  
 ; APPLICANT: HOFFMAN, No. 6309820h  
 ; APPLICANT: KAY, Brian K.  
 ; APPLICANT: FOWLKES, Dana M.  
 ; APPLICANT: McCONNELL, Stephen J.  
 ; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
 ; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
 ; TITLE OF INVENTION: USING SAME  
 ; NUMBER OF SEQUENCES: 227  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pennie & Edmonds LLP  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10036-2711

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 55 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-630-915A-111

```

```

Query Match          20.5%; Score 43.5; DB 4; Length 55;
Best Local Similarity 41.4%; Pred. No. 18;
Matches 12; Conservative 8; Mismatches 4; Indels 5; Gaps 2;

```

```

Qy      4 SISENSLVAMDFS-GQKSRVIENPTEALS 31
      :::: ||||: || ||::|   | | |:
Db      23 TVNKGSLVALGFSDGQEAR----PEEILN 47

```

```

RESULT 8
US-09-134-000C-5090
; Sequence 5090, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5090
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5090

```

```

Query Match          20.3%; Score 43; DB 4; Length 81;

```

Best Local Similarity 24.4%; Pred. No. 37;  
Matches 11; Conservative 11; Mismatches 13; Indels 10; Gaps 2;

Qy 9 SLVAMDFSGQKSRVIEN---PTEALSVAVEEGLA-----WRKK 43  
:| :|| :| | : : | :| :|| : :: | |:  
Db 7 ALEVIDFKSKDKRVNSKKIPPLKAIEVAKRKNVSAATVTRWMKR 51

RESULT 9

US-08-776-059-16

; Sequence 16, Application US/08776059B  
; Patent No. 6271368  
; GENERAL INFORMATION:  
; APPLICANT: LENTZEN, Hans  
; APPLICANT: ECK, Jorgen  
; APPLICANT: BAUR, Axel  
; APPLICANT: ZINKE, Holger  
; TITLE OF INVENTION: RECOMBINANT MISTLETOE LECTIN (RML)  
; FILE REFERENCE: 674503-2003  
; CURRENT APPLICATION NUMBER: US/08/776,059B  
; CURRENT FILING DATE: 1999-06-19  
; EARLIER APPLICATION NUMBER: PCT/EP96/02273  
; EARLIER FILING DATE: 1996-06-25  
; EARLIER APPLICATION NUMBER: 95109949.8  
; EARLIER FILING DATE: 1995-06-26  
; NUMBER OF SEQ ID NOS: 56  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 16  
; LENGTH: 47  
; TYPE: PRT  
; ORGANISM: Saponaria officinalis  
US-08-776-059-16

Query Match 19.8%; Score 42; DB 3; Length 47;  
Best Local Similarity 30.8%; Pred. No. 25;  
Matches 8; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

Qy 10 LVAMDFSGQKSRVIENPTEALSVAVE 35  
| |: :|:||:| | :|::  
Db 3 LTFMEAVNKKARVVKNEARFLLIAIQ 28

RESULT 10

US-09-134-000C-5679

; Sequence 5679, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 5679  
; LENGTH: 61  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-5679

Query Match 19.8%; Score 42; DB 4; Length 61;  
Best Local Similarity 34.1%; Pred. No. 36;  
Matches 15; Conservative 11; Mismatches 14; Indels 4; Gaps 3;

Qy 2 MRSISE--NSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
::|||| || : :: : : :|: |:| ||: || | |  
Db 12 LQSISEEPNSFI-IEETIKYIEQLEDDNESLQVAL-EGTIWSPK 53

RESULT 11

US-09-107-532A-5556

; Sequence 5556, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND

THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD-ROM ISO9660

; COMPUTER: PC

; OPERATING SYSTEM: <Unknown>

; SOFTWARE: ASCII

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,532A

; FILING DATE: 30-Jun-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/085,598

; FILING DATE: 14 May 1998

; APPLICATION NUMBER: 60/051571

; FILING DATE: July 2, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Ariniello, Pamela Deneke

; REGISTRATION NUMBER: 40,489

; REFERENCE/DOCKET NUMBER: GTC-012

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781)893-5007

; TELEFAX: (781)893-8277

; INFORMATION FOR SEQ ID NO: 5556:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 68 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

```

;      MOLECULE TYPE: protein
;      HYPOTHETICAL: YES
;      ORIGINAL SOURCE:
;          ORGANISM: Enterococcus faecium
;      FEATURE:
;          NAME/KEY:  misc_feature
;          LOCATION:  (B) LOCATION 1...68
;      SEQUENCE DESCRIPTION: SEQ ID NO: 5556:
US-09-107-532A-5556

```

```

Query Match          19.8%;  Score 42;  DB 4;  Length 68;
Best Local Similarity 20.9%;  Pred. No. 41;
Matches      9;  Conservative 14;  Mismatches 16;  Indels  4;  Gaps  1;

```

```

Qy      1 PMRSISENSLVAMDFSG----QKSRVIENPTEALSVAVEEGLA 39
        |: |  ::|:: |  :: : :  ||:: || :|::
Db      2 PLEDIRSIQIIAINIDGTLNNEERELTKEVKEAIAAAVAKGVS 44

```

#### RESULT 12

```

US-09-540-236-2395
; Sequence 2395, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
MORAXELLA CATARRHALIS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2395
; LENGTH: 78
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-2395

```

```

Query Match          19.8%;  Score 42;  DB 4;  Length 78;
Best Local Similarity 66.7%;  Pred. No. 50;
Matches      8;  Conservative 4;  Mismatches 0;  Indels  0;  Gaps  0;

```

```

Qy      28 EALSVAVEEGLA 39
        ||:|::|:|
Db      39 EAISVSLEDGLA 50

```

#### RESULT 13

```

US-09-198-452A-1167
; Sequence 1167, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides,
fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the
diagnosis, prevention

```

; TITLE OF INVENTION: and treatment of infection  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; CURRENT FILING DATE: 1998-11-24  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 1167  
; LENGTH: 81  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-1167

Query Match 19.8%; Score 42; DB 4; Length 81;  
Best Local Similarity 33.3%; Pred. No. 53;  
Matches 8; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

Qy 15 FSGQKSRVIENPTEALSVAVEEGL 38  
| |::||| |::| |:  
Db 30 FQGKRTRVIAITPAGLAIAYEQNI 53

RESULT 14

US-07-641-971B-5

; Sequence 5, Application US/07641971B  
; Patent No. 5236706  
; GENERAL INFORMATION:  
; APPLICANT: Debre, Patrice  
; APPLICANT: Mossalayi, Mohammed D  
; TITLE OF INVENTION: A PHARMACEUTICAL PREPARATION FOR THE  
; TITLE OF INVENTION: MATURATION OF PROTHYMOCYTES  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Irving M. Fishman, CIBA-GEIGY Corporation  
; STREET: 556 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/641,971B  
; FILING DATE: 19910116  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 90016254  
; FILING DATE: 24-JAN-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fishman, Irving M  
; REGISTRATION NUMBER: 30258  
; REFERENCE/DOCKET NUMBER: 4-17921/+/DEB  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 908-277-4832  
; TELEFAX: 908-277-4306  
; INFORMATION FOR SEQ ID NO: 5:



; SEQUENCE CHARACTERISTICS:  
; LENGTH: 40 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
US-07-641-971B-5

Query Match 19.3%; Score 41; DB 1; Length 40;  
Best Local Similarity 39.3%; Pred. No. 29;  
Matches 11; Conservative 5; Mismatches 8; Indels 4; Gaps 1;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTE 28  
|:|::: : |||| |: | |  
Db 2 PVRSLN----CTLRDSGQKSLVMSGPYE 25

RESULT 15

US-07-781-248A-5

; Sequence 5, Application US/07781248A  
; Patent No. 5246699

; GENERAL INFORMATION:

; APPLICANT: Debre, Patrice  
; APPLICANT: Mossalayi, Mohammed D  
; TITLE OF INVENTION: MATURATION OF HEMATOPOIETIC CELLS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Irving M. Fishman, CIBA-GEIGY Corporation  
; STREET: 556 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/781,248A  
; FILING DATE: 19911230  
; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 90103565  
; FILING DATE: 09-MAY-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Ikeler, Barbara J.  
; REGISTRATION NUMBER: 36,170  
; REFERENCE/DOCKET NUMBER: 4-18065/A/DEB

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 908-277-3368  
; TELEFAX: 908-277-4306

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 40 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
US-07-781-248A-5

Query Match 19.3%; Score 41; DB 1; Length 40;  
Best Local Similarity 39.3%; Pred. No. 29;  
Matches 11; Conservative 5; Mismatches 8; Indels 4; Gaps 1;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTE 28  
|:|::: : |||| |: | |  
Db 2 PVRSLN----CTLRDSGQKSLVMSGPYE 25

RESULT 16

US-09-107-532A-6894

; Sequence 6894, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND

THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD-ROM ISO9660

; COMPUTER: PC

; OPERATING SYSTEM: <Unknown>

; SOFTWARE: ASCII

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,532A

; FILING DATE: 30-Jun-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/085,598

; FILING DATE: 14 May 1998

; APPLICATION NUMBER: 60/051571

; FILING DATE: July 2, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Ariniello, Pamela Deneke

; REGISTRATION NUMBER: 40,489

; REFERENCE/DOCKET NUMBER: GTC-012

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781)893-5007

; TELEFAX: (781)893-8277

; INFORMATION FOR SEQ ID NO: 6894:

```

;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 61 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      HYPOTHETICAL: YES
;      ORIGINAL SOURCE:
;      ORGANISM: Enterococcus faecium
;      FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (B) LOCATION 1...61
;      SEQUENCE DESCRIPTION: SEQ ID NO: 6894:
US-09-107-532A-6894

```

```

Query Match          19.3%; Score 41; DB 4; Length 61;
Best Local Similarity 21.4%; Pred. No. 51;
Matches    9; Conservative 14; Mismatches 13; Indels 6; Gaps 1;

```

```

Qy      7 ENSLVAMDFSGQKSRVIENPTEALSVAV-----EEGLAWRK 42
      :::||:| : |::| ::: :: | ||:|
Db      1 KSEIVADGNANGSIILKNTPKSVQPSIFADSSKLSGKAWKK 42

```

#### RESULT 17

```

US-09-621-976-7338
; Sequence 7338, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 7338
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 78
; OTHER INFORMATION: Xaa = Asp,Glu
US-09-621-976-7338

```

```

Query Match          19.3%; Score 41; DB 4; Length 79;
Best Local Similarity 40.0%; Pred. No. 73;
Matches    10; Conservative 5; Mismatches 6; Indels 4; Gaps 1;

```

```

Qy      19 KSRVIENPTEALSVAVEEGLAWRKK 43
      |: :|| |:: | |:| |||
Db      49 KNEYVENRTKSR----EHGIAMRKK 69

```

#### RESULT 18

US-09-023-905A-17

; Sequence 17, Application US/09023905A  
; Patent No. 6475778  
; GENERAL INFORMATION:  
; APPLICANT: Roberts, Thomas M.  
; APPLICANT: King, Frederick J.  
; APPLICANT: Harris, David F.  
; APPLICANT: Hu, Erding  
; APPLICANT: Spiegelman, Bruce  
; APPLICANT: Chan, Joanne  
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses  
; TITLE OF INVENTION: Therefor  
; FILE REFERENCE: DFN-021  
; CURRENT APPLICATION NUMBER: US/09/023,905A  
; CURRENT FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/038,191  
; PRIOR FILING DATE: 1997-02-14  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 51  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-023-905A-17

Query Match 19.1%; Score 40.5; DB 4; Length 51;  
Best Local Similarity 47.4%; Pred. No. 48;  
Matches 9; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

Qy 4 SISENSLVAMDFS-GQKSR 21  
:::: ||||: || ||::|  
Db 22 TVNKGSLVALGFSDGQEAR 40

RESULT 19

US-08-459-568-52

; Sequence 52, Application US/08459568  
; Patent No. 5811304  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
; TITLE OF INVENTION: Zinc Finger Proteins  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/459,568

```

; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/399,411
; FILING DATE: 06-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-459-568-52

```

```

Query Match          19.1%; Score 40.5; DB 2; Length 66;
Best Local Similarity 47.4%; Pred. No. 68;
Matches      9; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

```

```

Qy      4 SISENSLVAMDFS-GQKSR 21
        :::: ||||: || ||::|
Db      22 TVNKGSLVALGFSDGQEAR 40

```

# RESULT 20

US-08-399-411-52

```

; Sequence 52, Application US/08399411
; Patent No. 5831008
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; TITLE OF INVENTION: Zinc Finger Proteins
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/399,411
; FILING DATE: 06-MAR-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1264

```

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 52:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 66 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-399-411-52

Query Match 19.1%; Score 40.5; DB 2; Length 66;  
Best Local Similarity 47.4%; Pred. No. 68;  
Matches 9; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

Qy 4 SISENSLVAMDFS-GQKSR 21  
::: |||: || ||:|  
Db 22 TVNKGSLVALGFSDGQEAR 40

RESULT 21  
US-08-516-859A-52  
; Sequence 52, Application US/08516859A  
; Patent No. 6069231  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
; TITLE OF INVENTION: Zinc Finger Proteins  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/516,859A  
; FILING DATE: 18-AUG-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/399,411  
; FILING DATE: 06-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/292,683  
; FILING DATE: 18-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 1776  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949

; INFORMATION FOR SEQ ID NO: 52:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 66 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-516-859A-52

Query Match 19.1%; Score 40.5; DB 3; Length 66;  
Best Local Similarity 47.4%; Pred. No. 68;  
Matches 9; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

Qy 4 SISENSLVAMDFS-GQKSR 21  
::: ||||: || ||::|  
Db 22 TVNKGSLVALGFSDGQEAR 40

RESULT 22

US-09-586-472-52

; Sequence 52, Application US/09586472  
; Patent No. 6323335  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
; Zinc Finger Proteins  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/586,472  
; FILING DATE: 01-Jun-2000  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/528,706  
; FILING DATE: 17-MAR-2000  
; APPLICATION NUMBER: US 08/516,859  
; FILING DATE: 18-AUG-1995  
; APPLICATION NUMBER: US 08/399,411  
; FILING DATE: 06-MAR-1995  
; APPLICATION NUMBER: US 08/292,683  
; FILING DATE: 18-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 4130  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949

; INFORMATION FOR SEQ ID NO: 52:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 66 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 52:  
US-09-586-472-52

Query Match 19.1%; Score 40.5; DB 4; Length 66;  
Best Local Similarity 47.4%; Pred. No. 68;  
Matches 9; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

Qy 4 SISENSLVAMDFS-GQKSR 21  
::: ||||: || ||::|  
Db 22 TVNKGSLVALGFSDGQEAR 40

RESULT 23

US-09-528-706-52

; Sequence 52, Application US/09528706  
; Patent No. 6468985  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
; TITLE OF INVENTION: Zinc Finger Proteins  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/528,706  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/516,859  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/292,683  
; FILING DATE: 18-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 1776  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 52:  
; SEQUENCE CHARACTERISTICS:



; LENGTH: 66 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-09-528-706-52

Query Match 19.1%; Score 40.5; DB 4; Length 66;  
Best Local Similarity 47.4%; Pred. No. 68;  
Matches 9; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

Qy 4 SISENSLVAMDFS-GQKSR 21  
::: |||: || ||::|  
Db 22 TVNKGSLVALGFSDGQEAR 40

RESULT 24

US-09-621-976-5251  
; Sequence 5251, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 5251  
; LENGTH: 71  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -21..-1  
US-09-621-976-5251

Query Match 19.1%; Score 40.5; DB 4; Length 71;  
Best Local Similarity 29.5%; Pred. No. 75;  
Matches 13; Conservative 6; Mismatches 22; Indels 3; Gaps 2;

Qy 2 MRSISEN--SLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
||::| | : |::| | | :| | | |  
Db 1 MRNLSSNLHGLCLLLLCQATGRIMEKTTH-LFFTCKENLGWNSK 43

RESULT 25

US-09-006-428A-14  
; Sequence 14, Application US/09006428A  
; Patent No. 6444439  
; GENERAL INFORMATION:  
; APPLICANT: Jing Li  
; APPLICANT: Kazuhisa Nishizawa  
; APPLICANT: Wenqian An  
; APPLICANT: Ellis L. Reinherz  
; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF A  
; TITLE OF INVENTION: cdc15-LIKE ADAPTOR PROTEIN (CD2BP1)

; FILE REFERENCE: 1062.1020-000  
; CURRENT APPLICATION NUMBER: US/09/006,428A  
; CURRENT FILING DATE: 1998-01-13  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 79  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-006-428A-14

Query Match 19.1%; Score 40.5; DB 4; Length 79;  
Best Local Similarity 47.4%; Pred. No. 87;  
Matches 9; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

Qy 4 SISENSLVAMDFS-GQKSR 21  
::: ||||: || ||::|  
Db 31 TVNKGSLVALGFSDGQEAR 49

Search completed: July 8, 2004, 08:23:30  
Job time : 12.8346 secs

OM protein - protein search, using sw model

Run on: July 8, 2004, 08:06:23 ; Search time 9.14173 Seconds  
 (without alignments)  
 452.456 Million cell updates/sec

Title: US-09-936-697-5  
 Perfect score: 212  
 Sequence: 1 PMRSISENSLVAMDFSGQKS.....ENPTEALSVAVEEGLAWRKK 43

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 28653

Minimum DB seq length: 0  
 Maximum DB seq length: 85

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 100 summaries

Database : PIR\_78:\*  
 1: pir1:\*  
 2: pir2:\*  
 3: pir3:\*  
 4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	% Query		DB	ID	Description
	Score	Match Length			
1	46.5	21.9	62	2	E64510
2	46	21.7	76	2	E64324
3	45.5	21.5	73	2	T25763
4	45	21.2	60	2	A42960
5	45	21.2	77	2	H69420
6	45	21.2	82	2	D69087
7	42	19.8	53	2	C82776
8	42	19.8	79	2	A86517
9	42	19.8	79	2	B72106
10	42	19.8	81	2	D81565
11	42	19.8	82	1	G64370
12	40.5	19.1	73	2	A60172
13	40	18.9	66	2	A43602

14	40	18.9	67	2	G97092	endoglucanase (tru
15	40	18.9	70	2	E70985	hypothetical prote
16	39.5	18.6	61	2	F86696	4-oxalocrotonate t
17	39.5	18.6	67	2	AF1487	probable transcrip
18	39.5	18.6	77	2	AD1945	hypothetical prote
19	39	18.4	45	1	C64901	ribosomal protein
20	39	18.4	45	2	D90889	30S ribosomal subu
21	39	18.4	45	2	E85728	30S ribosomal subu
22	39	18.4	62	2	T06654	hypothetical prote
23	39	18.4	72	2	AD3532	hypothetical prote
24	39	18.4	77	2	B83269	hypothetical prote
25	38.5	18.2	48	2	T35253	small hypothetical
26	38.5	18.2	67	2	T42055	cold shock protein
27	38.5	18.2	67	2	AH1375	repressor protein
28	38.5	18.2	67	2	AI1126	probable transcrip
29	38.5	18.2	69	2	S70158	hypothetical prote
30	38.5	18.2	78	1	EWSMYG	cinnamycin precurs
31	38	17.9	59	1	FEMZB	ferredoxin 2[4Fe-4
32	38	17.9	63	1	FEDV2V	ferredoxin [4Fe-4S
33	38	17.9	64	2	D97731	hypothetical prote
34	38	17.9	74	2	T26580	hypothetical prote
35	38	17.9	78	2	PS0085	hypothetical prote
36	37.5	17.7	68	2	T00189	hypothetical prote
37	37.5	17.7	74	2	T48037	hypothetical prote
38	37.5	17.7	83	2	A82386	hypothetical prote
39	37	17.5	66	2	AH0343	hypothetical prote
40	37	17.5	71	2	D98035	hypothetical prote
41	37	17.5	84	2	S33432	hypothetical prote
42	36.5	17.2	83	2	A89940	30S ribosomal prot
43	36.5	17.2	84	2	E69293	conserved hypothet
44	36	17.0	34	4	JT0745	hypothetical trans
45	36	17.0	36	2	JC2006	differentiation in
46	36	17.0	52	2	A34657	cathepsin E (EC 3.
47	36	17.0	56	2	C72610	hypothetical prote
48	36	17.0	59	2	H69463	hypothetical prote
49	36	17.0	65	2	B95398	protein [imported
50	36	17.0	66	2	S26378	cold shock protein
51	36	17.0	68	2	S20920	1-aminocyclopropan
52	36	17.0	71	2	B89990	hypothetical prote
53	36	17.0	74	2	AH3104	conserved hypothet
54	36	17.0	75	2	AC3575	hypothetical prote
55	36	17.0	78	2	C98182	hypothetical prote
56	36	17.0	80	2	S02067	glycerol kinase (E
57	36	17.0	84	2	G69079	conserved hypothet
58	35.5	16.7	62	2	C71125	hypothetical prote
59	35.5	16.7	69	2	A56378	GTP-binding regula
60	35.5	16.7	75	2	S07730	H <sup>+</sup> -transporting tw
61	35.5	16.7	75	2	A70610	hypothetical prote
62	35.5	16.7	78	2	H90114	putative small nuc
63	35	16.5	53	2	S56726	calcium-dependent
64	35	16.5	58	2	A45824	hypothetical prote
65	35	16.5	61	2	A81781	probable membrane
66	35	16.5	63	2	B55879	cannabinoid recept
67	35	16.5	63	2	AH2565	hypothetical prote
68	35	16.5	70	1	EILXCH	eglin C - medicina
69	35	16.5	74	2	T03188	hypothetical prote
70	35	16.5	77	2	AC3626	glutaredoxin [impo

71	35	16.5	78	2	D81246	hypothetical prote
72	35	16.5	81	2	D97063	uncharacterized pr
73	35	16.5	85	2	AG3408	molybdopterin (mpt
74	34.5	16.3	55	2	D82595	hypothetical prote
75	34.5	16.3	58	2	S30499	protein-tyrosine k
76	34.5	16.3	61	2	A35567	photosystem I 14K
77	34.5	16.3	71	2	E98040	degenerative trans
78	34.5	16.3	77	2	B95003	hypothetical prote
79	34.5	16.3	77	2	B98247	hypothetical prote
80	34.5	16.3	81	2	T01671	vpu protein - huma
81	34.5	16.3	81	2	E59093	hypothetical prote
82	34	16.0	39	2	F95042	hypothetical prote
83	34	16.0	42	2	I65746	tropomyosin - huma
84	34	16.0	51	2	E72807	gp62 protein - Myc
85	34	16.0	56	2	AB2413	hypothetical prote
86	34	16.0	61	2	F84126	4-oxalocrotonate t
87	34	16.0	64	2	H98033	hypothetical prote
88	34	16.0	65	2	T03652	probable carbonate
89	34	16.0	66	2	G82491	hypothetical prote
90	34	16.0	66	2	T12847	hypothetical prote
91	34	16.0	68	2	JQ0099	hypothetical 7K pr
92	34	16.0	68	2	G69209	histone-related pr
93	34	16.0	69	2	AB1333	hypothetical prote
94	34	16.0	69	2	AB1704	hypothetical prote
95	34	16.0	71	2	AF2303	hypothetical prote
96	34	16.0	72	2	S54448	hypothetical prote
97	34	16.0	74	2	F96670	hypothetical prote
98	34	16.0	78	2	S03233	hypothetical prote
99	34	16.0	79	2	S03448	T-cell receptor al
100	34	16.0	82	2	G81247	BolA/YrbA family p

#### ALIGNMENTS

##### RESULT 1

E64510

hypothetical protein MJECL05 - Methanococcus jannaschii plasmid pURB800

C;Species: Methanococcus jannaschii

C;Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 22-Oct-1999

C;Accession: E64510

R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghagen, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.

A;Reference number: A64300; MUID:96337999; PMID:8688087

A;Accession: E64510

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-62 <BUL>  
A;Cross-references: GB:L77118; NID:g1500644; TIGR:MJECL05; PIDN:AAC37071.1;  
PID:g1500645  
C;Genetics:  
A;Map position: ECLFOR3265-3453  
A;Genome: plasmid  
A;Start codon: GTG  
A;Note: this stable 58-kilobase pair plasmid is also designated ECL (large  
extrachromosomal element) and contains 44 predicted coding regions

Query Match 21.9%; Score 46.5; DB 2; Length 62;  
Best Local Similarity 28.6%; Pred. No. 28;  
Matches 12; Conservative 8; Mismatches 21; Indels 1; Gaps 1;

Qy 3 RSISENSLVAMDFS-GQKSRVIENPTEALSVAVEEGLAWRKK 43  
: : : | | : : | | : | | : | | : | | : | |  
Db 18 KKVAERFLKDLESSQGMDWKEIRERAERAKKQLEEGIEWAKK 59

# RESULT 2 E64324

DNA-directed RNA polymerase (EC 2.7.7.6) subunit N - Methanococcus jannaschii  
C;Species: Methanococcus jannaschii  
C;Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 23-Apr-1999  
C;Accession: E64324  
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.;  
Blake, J.A.; FitzGerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.;  
Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness,  
E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghagen,  
N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley,  
J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.;  
Hurst, M.A.  
Science 273, 1058-1073, 1996  
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.;  
Woese, C.R.; Venter, J.C.  
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus  
jannaschii.  
A;Reference number: A64300; MUID:96337999; PMID:8688087  
A;Accession: E64324  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-76 <BUL>  
A;Cross-references: GB:U67475; GB:L77117; NID:g1590930; PID:g1590941;  
TIGR:MJ0196; PID:g1510312  
C;Genetics:  
A;Map position: FOR190573-190803  
A;Start codon: GTG  
C;Superfamily: DNA-directed RNA polymerase II chain RPB10  
C;Keywords: nucleotidyltransferase; transcription

Query Match 21.7%; Score 46; DB 2; Length 76;  
Best Local Similarity 34.4%; Pred. No. 41;  
Matches 11; Conservative 7; Mismatches 12; Indels 2; Gaps 1;

Qy 1 PMRSISENSLVAMDFSGQKSRVI--ENPTEAL 30  
| : | | : : | | | : | | : | | : |  
Db 7 PIRCFSCGNVIAEVFEEYKERILKGENPKDVL 38

# RESULT 3

T25763

hypothetical protein F46F11.4 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T25763

R;Pauley, A.; Gattung, S.

submitted to the EMBL Data Library, February 1997

A;Description: The sequence of *C. elegans* cosmid F46F11.

A;Reference number: Z20083

A;Accession: T25763

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-73 <PAU>

A;Cross-references: EMBL:U88173; PIDN:AAB42266.1; GSPDB:GN00019; CESP:F46F11.4

A;Experimental source: strain Bristol N2; clone F46F11

C;Genetics:

A;Gene: CESP:F46F11.4

A;Map position: 1

A;Introns: 38/2

Query Match 21.5%; Score 45.5; DB 2; Length 73;  
Best Local Similarity 29.4%; Pred. No. 46;  
Matches 10; Conservative 7; Mismatches 12; Indels 5; Gaps 1;

Qy 14 DFGGQKSRVIENPTEALS-----VAVEEGLAWRK 42

| |:| |: ||:: : :| :| ||

Db 8 DRLGKKVRIKCNPSDTIGDLKKLIAAQTGTRWEK 41

# RESULT 4

A42960

ferredoxin 2[4Fe-4S] - *Methanosarcina thermophila*

C;Species: *Methanosarcina thermophila*

C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 13-Nov-1998

C;Accession: A42960

R;Clements, A.P.; Ferry, J.G.

J. Bacteriol. 174, 5244-5250, 1992

A;Title: Cloning, nucleotide sequence, and transcriptional analyses of the gene encoding a ferredoxin from *Methanosarcina thermophila*.

A;Reference number: A42960; MUID:92355496; PMID:1379583

A;Contents: TM-1

A;Accession: A42960

A;Molecule type: DNA

A;Residues: 1-60 <CLE>

A;Note: sequence extracted from NCBI backbone (NCBIN:110322, NCBIP:110324)

C;Genetics:

A;Gene: fdxA

C;Superfamily: ferredoxin 2[4Fe-4S]; ferredoxin 2[4Fe-4S] homology

C;Keywords: 4Fe-4S; electron transfer; iron-sulfur protein; metalloprotein

F;3-59/Domain: ferredoxin 2[4Fe-4S] homology <FER>

F;10,13,16,51/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

F;20,41,44,47/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 21.2%; Score 45; DB 2; Length 60;

Best Local Similarity 42.9%; Pred. No. 43;  
Matches 12; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

Qy 12 AMDFSGQKSRVIENPTEALSVAVEEGLA 39  
| : || | | | :||::: |:|:  
Db 7 ADECSGCGSCVDECPSEAITLDEEKGIA 34

RESULT 5

H69420

hydrogenase expression/formation protein (hypC) homolog - *Archaeoglobus fulgidus*  
C;Species: *Archaeoglobus fulgidus*

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 14-Apr-2003

C;Accession: H69420

R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*.

A;Reference number: A69250; MUID:98049343; PMID:9389475

A;Accession: H69420

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-77 <KLE>

A;Cross-references: GB:AE001009; GB:AE000782; NID:g2689332; PIDN:AAB89878.1; PID:g2649207; TIGR:AF1369

C;Superfamily: [NiFe]-hydrogenase maturation chaperone

Query Match 21.2%; Score 45; DB 2; Length 77;  
Best Local Similarity 35.1%; Pred. No. 57;  
Matches 13; Conservative 5; Mismatches 15; Indels 4; Gaps 1;

Qy 10 LVAMDFSGQKSRV----IENPTEALSVAVEEGLAWRK 42  
: :|| | | | :|| | | |:| :|  
Db 16 IAIVDFKGLKKEVRIDLLENPQIGDYVLVHVGMAIQK 52

RESULT 6

D69087

hydrogenase expression/formation protein HypC - *Methanobacterium thermoautotrophicum* (strain Delta H)

C;Species: *Methanobacterium thermoautotrophicum*

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 14-Apr-2003

C;Accession: D69087

R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.;



Pietrokovski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.;  
Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A;Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta  
H: functional analysis and comparative genomics.

A;Reference number: A69000; MUID:98037514; PMID:9371463

A;Accession: D69087

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-82 <MTH>

A;Cross-references: GB:AE000924; GB:AE000666; NID:g2622777; PIDN:AAB86122.1;

PID:g2622778

A;Experimental source: strain Delta H

C;Genetics:

A;Gene: MTH1649

C;Superfamily: [NiFe]-hydrogenase maturation chaperone

Query Match 21.2%; Score 45; DB 2; Length 82;

Best Local Similarity 28.9%; Pred. No. 62;

Matches 11; Conservative 9; Mismatches 14; Indels 4; Gaps 1;

Qy 6 SENSLVAMDFSGQKSRV----IENPTEALSVAVEEGLA 39

||::: || | : | :: | | | |

Db 14 SEDNIATVDFGGVRQQVKLDLVDDVEEGKYVLVHSGYA 51

#### RESULT 7

C82776

hypothetical protein XF0694 [imported] - *Xylella fastidiosa* (strain 9a5c)

C;Species: *Xylella fastidiosa*

C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C;Accession: C82776

R;anonymous, The *Xylella fastidiosa* Consortium of the Organization for  
Nucleotide Sequencing and Analysis, Sao Paulo, Brazil.

Nature 406, 151-157, 2000

A;Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A;Reference number: A82515; MUID:20365717; PMID:10910347

A;Note: for a complete list of authors see reference number A59328 below

A;Accession: C82776

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-53 <SIM>

A;Cross-references: GB:AE003912; GB:AE003849; NID:g9105560; PIDN:AAF83504.1;

GSPDB:GN00128; XFSC:XF0694

A;Experimental source: strain 9a5c

R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.;

Alvarenga, R.; Alves, L.M.C.; Araya, J.E.; Baia, G.S.; Baptista, C.S.; Barros,

M.H.; Bonaccorsi, E.D.; Bordin, S.; Bove, J.M.; Briones, M.R.S.; Bueno, M.R.P.;

Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.; Colauto, N.B.;

Colombo, C.; Costa, F.F.; Costa, M.C.R.; Costa-Neto, C.M.; Coutinho, L.L.;

Cristofani, M.; Dias-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.;

Ferreira, A.J.S.

submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco,

M.C.; Frohme, M.; Furlan, L.R.; Garnier, M.; Goldman, G.H.; Goldman, M.H.S.;

Gomes, S.L.; Gruber, A.; Ho, P.L.; Hoheisel, J.D.; Junqueira, M.L.; Kemper,

E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret, F.; Lambais, M.R.;

Leite, L.C.C.; Lemos, E.G.M.; Lemos, M.V.F.; Lopes, S.A.; Lopes, C.R.; Machado, J.A.; Machado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.L.

A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; Monteiro-Vitorello, C.B.; Moon, D.H.; Nagai, M.A.; Nascimento, A.L.T.O.; Netto, L.E.S.; Nhani Jr., A.; Nobrega, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Paris, A.; Peixoto, B.R.; Pereira, G.A.G.; Pereira Jr., H.A.; Pesquero, J.B.; Quaggio, R.B.; Roberto, P.G.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki, H.E.

A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, J.F.; Silvestri, M.L.Z.; Siqueira, W.J.; de Souza, A.A.; de Souza, A.P.; Terenzi, M.F.; Truffi, D.; Tsai, S.M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Zago, M.A.; Zatz, M.; Meidanis, J.; Setubal, J.C.

A;Reference number: A59328

A;Contents: annotation

C;Genetics:

A;Gene: XF0694

Query Match 19.8%; Score 42; DB 2; Length 53;  
Best Local Similarity 66.7%; Pred. No. 95;  
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 30 LSVAVEEGLAWR 41  
| | || | |||  
Db 21 LGVGVERGYAWR 32

#### RESULT 8

A86517

hypothetical protein CPj0209 [imported] - Chlamydophila pneumoniae (strain J138)

C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001

C;Accession: A86517

R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ishii, K.; Hattori, M.; Kuhara, S.; Nakazawa, T.

Nucleic Acids Res. 28, 2311-2314, 2000

A;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.

A;Reference number: A86491; MUID:20330349; PMID:10871362

A;Accession: A86517

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-79 <STO>

A;Cross-references: GB:BA000008; NID:g8978582; PIDN:BAA98419.1; GSPDB:GN00142

A;Experimental source: strain J138

C;Genetics:

A;Gene: CPj0209

Query Match 19.8%; Score 42; DB 2; Length 79;  
Best Local Similarity 33.3%; Pred. No. 1.5e+02;  
Matches 8; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

Qy 15 FSGQKSRVIENPTEALSVAVEEGL 38  
| |::||| |::| |::  
Db 28 FQKRTRVIAITPAGLAIAEQNI 51

RESULT 9

B72106

hypothetical protein - Chlamydophila pneumoniae (strain CWL029)

C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae

C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000

C;Accession: B72106

R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.;

Grimwood, J.; Davis, R.W.; Stephens, R.S.

Nature Genet. 21, 385-389, 1999

A;Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A;Reference number: A72000; MUID:99206606; PMID:10192388

A;Accession: B72106

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-79 <ARN>

A;Cross-references: GB:AE001607; GB:AE001363; NID:g4376474; PIDN:AAD18362.1; PID:g4376481

A;Experimental source: strain CWL029

C;Genetics:

A;Gene: CPn0209

Query Match 19.8%; Score 42; DB 2; Length 79;  
Best Local Similarity 33.3%; Pred. No. 1.5e+02;  
Matches 8; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

Qy 15 FSGQKSRVIENPTEALSVAVEEGL 38

| |::||| |::| |: :

Db 28 FQGKRTRVIAITPAGLAIAYEQNI 51

RESULT 10

D81565

hypothetical protein CP0557 [imported] - Chlamydophila pneumoniae (strain AR39)

C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae

C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000

C;Accession: D81565

R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.;

Hickey, E.K.; Peterson, J.; Utterback, T.; Berry, K.; Bass, S.; Linher, K.;

Weidman, J.; Khouri, H.; Craven, B.; Bowman, C.; Dodson, R.; Gwinn, M.; Nelson,

W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, S.L.; Eisen, J.; Fraser,

C.M.

Nucleic Acids Res. 28, 1397-1406, 2000

A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A;Reference number: A81500; MUID:20150255; PMID:10684935

A;Accession: D81565

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-81 <REA>

A;Cross-references: GB:AE002214; GB:AE002161; NID:g7189460; PIDN:AAF38377.1;

PID:g7189469; GSPDB:GN00122; TIGR:CP0557

A;Experimental source: strain AR39, HL cells

C;Genetics:

A;Gene: CP0557

Query Match 19.8%; Score 42; DB 2; Length 81;

Best Local Similarity 33.3%; Pred. No. 1.6e+02;  
Matches 8; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

Qy 15 FSGQKSRVIENPTEALSVAVEEGL 38  
| |::| | | |::| | : :  
Db 30 FQKRTTRVIAITPAGLAIAYEQNI 53

RESULT 11

G64370

conserved hypothetical protein MJ0567 - *Methanococcus jannaschii*

C;Species: *Methanococcus jannaschii*

C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000

C;Accession: G64370

R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.;  
Blake, J.A.; FitzGerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.;  
Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness,  
E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghagen,  
N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley,  
J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.;  
Hurst, M.A.

Science 273, 1058-1073, 1996

A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.;  
Woese, C.R.; Venter, J.C.

A;Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*.

A;Reference number: A64300; MUID:96337999; PMID:8688087

A;Accession: G64370

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-82 <BUL>

A;Cross-references: GB:U67505; GB:L77117; NID:g2826297; PIDN:AAB98558.1;  
PID:g1591273; TIGR:MJ0567

C;Genetics:

A;Map position: REV504744-504496

C;Superfamily: *Methanococcus jannaschii* conserved hypothetical protein MJ0567

Query Match 19.8%; Score 42; DB 1; Length 82;  
Best Local Similarity 32.5%; Pred. No. 1.6e+02;  
Matches 13; Conservative 8; Mismatches 11; Indels 8; Gaps 2;

Qy 10 LVAMDFS-GQKSRVIEN-----PTEALSVAVEEGLAWR 41  
|::| : | | :|| | |::| :|| :  
Db 28 LVSMGINIGSKLKVIRNQNGPVIISTKGSNIAIGRGLAMK 67

RESULT 12

A60172

proteoglycan core protein, laryngeal cartilage - pig (fragments)

C;Species: *Sus scrofa domestica* (domestic pig)

C;Date: 22-Jan-1993 #sequence\_revision 22-Jan-1993 #text\_change 13-Aug-1999

C;Accession: A60172

R;Harris, M.; Kenneally, B.; Barry, F.

Biochem. Soc. Trans. 18, 299, 1990

A;Title: Primary structure of the hyaluronic acid-binding region of porcine laryngeal cartilage proteoglycan.

A;Reference number: A60172; MUID:90337042; PMID:1696222

A;Accession: A60172  
A;Molecule type: protein  
A;Residues: 1-73 <HAR>  
C;Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EGF homology; immunoglobulin homology; link protein repeat homology  
C;Keywords: cartilage  
F;41-73/Domain: link protein repeat homology (fragment) <LNK1>

Query Match 19.1%; Score 40.5; DB 2; Length 73;  
Best Local Similarity 38.7%; Pred. No. 2.2e+02;  
Matches 12; Conservative 3; Mismatches 7; Indels 9; Gaps 2;

Qy 18 QKSRVIENPTEALSVAVEEG-----LAW 40  
| | :| | | :| | :| | |  
Db 44 QNSAIIATP-ENLNAATEDGPHQCDAGWLAW 73

#### RESULT 13

A43602  
T-cell-stimulating antigen - *Coccidioides immitis* (fragment)  
C;Species: *Coccidioides immitis*  
C;Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 11-Jan-2000  
C;Accession: A43602; S16764  
R;Kirkland, T.N.; Zhu, S.; Kruse, D.; Hsu, L.; Seshan, K.R.; Cole, G.T.  
Infect. Immun. 59, 3952-3961, 1991  
A;Title: *Coccidioides immitis* fractions which are antigenic for immune T lymphocytes.  
A;Reference number: A43602; MUID:92040063; PMID:1840578  
A;Accession: A43602  
A;Molecule type: mRNA  
A;Residues: 1-66 <KIR>  
A;Cross-references: GB:M77190  
A;Note: the authors translated the codon ACC for residue 61 as Asn  
C;Superfamily: human 4-hydroxyphenylpyruvate dioxygenase

Query Match 18.9%; Score 40; DB 2; Length 66;  
Best Local Similarity 26.7%; Pred. No. 2.3e+02;  
Matches 8; Conservative 9; Mismatches 13; Indels 0; Gaps 0;

Qy 14 DFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
:| | || | :| : : : | :| :|  
Db 1 EFSALKSIVMASPNDIVKMPINEPAKGKKQ 30

#### RESULT 14

G97092  
endoglucanase (truncated) [imported] - *Clostridium acetobutylicum*  
C;Species: *Clostridium acetobutylicum*  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C;Accession: G97092  
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, H.M.; Dubois, J.; Qiu, D.; Hitti, J.; Wolf, Y.I.; Tatusov, R.L.; Sabathe, F.; Doucette-Stamm, L.; Soucaille, P.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum*.

A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: G97092  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-67 <KUR>  
A;Cross-references: GB:AE001437; PIDN:AAK79530.1; PID:g15024515; GSPDB:GN00168  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC1563

RESULT 15

hypothetical protein Rv1740 - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Nov-2000  
C;Accession: E70985  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.;  
Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry III, C.E.; Tekaiia, F.; Badcock, K.;  
Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;  
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.;  
Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail,  
M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.;  
Squares, S.  
Nature 393, 537-544, 1998  
A;Authors: Sqaers, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete  
genome sequence.  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: E70985  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-70 <COL>  
A;Cross-references: GB:Z95890; GB:AL123456; NID:g3242245; PIDN:CAB09326.1;  
PID:e318980; PID:g2131004  
A;Experimental source: strain H37Rv  
C;Genetics:  
A;Gene: Rv1740  
C;Superfamily: Mycobacterium tuberculosis hypothetical protein Rv0608

RESULT 16

F86696

4-oxalocrotonate tautomerase [imported] - *Lactococcus lactis* subsp. *lactis*  
(strain IL1403)

C;Species: *Lactococcus lactis* subsp. *lactis*

C;Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001

C;Accession: F86696

R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich, S.D.; Sorokin, A.

Genome Res. 11, 731-753, 2001

A;Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403.

A;Reference number: A86625; MUID:21235186; PMID:11337471

A;Accession: F86696

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-61 <STO>

A;Cross-references: GB:AE005176; PID:g12723464; PIDN:AAK04672.1; GSPDB:GN00146

A;Experimental source: strain IL1403

C;Genetics:

A;Gene: xylH

Query Match 18.6%; Score 39.5; DB 2; Length 61;  
Best Local Similarity 21.4%; Pred. No. 2.4e+02;  
Matches 9; Conservative 14; Mismatches 16; Indels 3; Gaps 1;

Qy 3 RSISENSLVAMDFSGQKSRVIENPTEALSVA---VEEGLAWR 41

|:: : ::| : : | : || | : | : ||: ::

Db 12 RTVEQKAIIAKEITESISKHAGAPTSAIHVIFNDLPEGMLYQ 53

RESULT 17

AF1487

probable transcription regulator homolog lin0437 [imported] - *Listeria innocua*  
(strain Clip11262)

C;Species: *Listeria innocua*

C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C;Accession: AF1487

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, H.; Brandt, P.; Chakraborty, T.; Charbit, A.; Chetouani, F.; Couve, E.; de Daruvar, A.; Dehoux, P.; Domann, E.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Garcia-Del Portillo, F.; Garrido, P.; Gautier, L.; Goebel, W.; Gomez-Lopez, N.; Hain, T.; Hauf, J.; Jackson, D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mata Vicente, J.; Ng, E.; Nordsiek, G.; Novella, S.; de Pablos, B.; Perez-Diaz, J.C.; Remmel, B.; Rose, M.; Rusniok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, J.; Cossart, P.

A;Title: Comparative genomics of *Listeria* species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AF1487

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-67 <GLA>

A;Cross-references: GB:AL592022; PIDN:CAC95670.1; PID:g16412866; GSPDB:GN00178

A;Experimental source: strain Clip11262

C;Genetics:  
A;Gene: lin0437

Query Match 18.6%; Score 39.5; DB 2; Length 67;  
Best Local Similarity 37.1%; Pred. No. 2.7e+02;  
Matches 13; Conservative 7; Mismatches 10; Indels 5; Gaps 2;

Qy 3 RSISENSL-VAMDFSGQKSRVIE----NPTEALSV 32  
|:| :| | :|:: | | || ||: ||:  
Db 14 RAIGQNELALALEVSRQTIHAIEKGKYNPSLELSL 48

RESULT 18

AD1945

hypothetical protein asl1111 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002

C;Accession: AD1945

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.;  
Iriguchi, M.; Ishikawa, A.; Kawashima, K.; Kimura, T.; Kishida, Y.; Kohara, M.;  
Matsumoto, M.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.;  
Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing  
Cyanobacterium Anabaena sp. strain PCC 7120.

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AD1945

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-77 <KUR>

A;Cross-references: GB:BA000019; PIDN:BAB73068.1; PID:g17130457; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: asl1111

Query Match 18.6%; Score 39.5; DB 2; Length 77;  
Best Local Similarity 22.2%; Pred. No. 3.2e+02;  
Matches 8; Conservative 10; Mismatches 13; Indels 5; Gaps 1;

Qy 2 MRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEG 37  
::: || : :: : || : :| |||  
Db 10 LKAVKENQALR-----ERLQATNNPEAFIKIAQEEG 40

RESULT 19

C64901

ribosomal protein S22 [validated] - Escherichia coli (strain K-12)

C;Species: Escherichia coli

C;Date: 24-Sep-1999 #sequence\_revision 24-Sep-1999 #text\_change 01-Mar-2002

C;Accession: C64901

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;  
Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,  
J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503



A;Accession: C64901  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-45 <BLAT>  
 A;Cross-references: GB:AE000245; GB:U00096; NID:g1787752; PIDN:AAC74553.1; PID:g1787755; UWGP:b1480  
 A;Experimental source: strain K-12, substrain MG1655  
 R;Arnold, R.J.; Reilly, J.P.  
 Anal. Biochem. 269, 105-112, 1999  
 A;Title: Observation of Escherichia coli ribosomal proteins and their posttranslational modifications by mass spectrometry.  
 A;Reference number: A59071; MUID:99196679; PMID:10094780  
 A;Contents: annotation; mass spectrographic analysis  
 A;Note: a ribosomal protein with these mass spectrographic characteristics was found; no post-translational modifications were observed in mass spectrographic analysis; any acid labile modifications may have been missed  
 C;Genetics:  
 A;Gene: rpsV  
 C;Complex: the ribosome is composed of the large (50S) and small (30S) subunit; the large (50S) subunit consists of 23S rRNA, 5S rRNA, and 34 distinct proteins; the small (30S) subunit consists of 16S rRNA and 22 distinct proteins  
 C;Complex: small subunit ribosomal proteins: S1 (PIR:R3EC1), S2 (PIR:R3EC2), S3 (PIR:R3EC3), S4 (PIR:R3EC4), S5 (PIR:R3EC5), S6 (PIR:R3EC6), S7 (PIR:R3EC7K), S8 (PIR:R3EC8), S9 (PIR:R3EC9), S10 (PIR:R3EC10), S11 (PIR:R3EC11), S12 (PIR:R3EC12), S13 (PIR:R3EC13), S14 (PIR:R3EC14), S15 (PIR:R3EC15), S16 (PIR:R3EC16), S17 (PIR:R3EC17), S18 (PIR:R3EC18), S19 (PIR:R3EC19), S20/L26 (PIR:R3EC20), S21 (PIR:R3EC21), S22 (PIR:C64901) [validated, MUID:99196679]  
 C;Function:  
 A;Pathway: protein biosynthesis  
 C;Superfamily: Escherichia coli ribosomal protein S22  
 C;Keywords: protein biosynthesis; ribosome  
 F;1-45/Product: ribosomal protein S22 #status experimental <MAT>

Query Match 18.4%; Score 39; DB 1; Length 45;  
 Best Local Similarity 63.6%; Pred. No. 2e+02;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 17 GQKSRVIENPT 27  
 | || |: |||  
 Db 27 GDKSSVVNNPT 37

## RESULT 20

D90889

30S ribosomal subunit protein S22 [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)

C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 17-May-2002

C;Accession: D90889

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Ohtsubo, E.; Nakayama, K.; Murata, T.; Tanaka, M.; Tobe, T.; Iida, T.; Takami, H.; Honda, T.; Sasakawa, C.; Ogasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.

A;Reference number: A99629; MUID:21156231; PMID:11258796

A;Accession: D90889  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-45 <HAY>  
A;Cross-references: GB:BA000007; PIDN:BAB35507.1; PID:g13361550; GSPDB:GN00154  
A;Experimental source: strain O157:H7, substrain RIMD 0509952  
C;Genetics:  
A;Gene: ECs2084  
C;Superfamily: Escherichia coli ribosomal protein S22

Query Match 18.4%; Score 39; DB 2; Length 45;  
Best Local Similarity 63.6%; Pred. No. 2e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 17 GQKSRVIENPT 27  
| || |: |||  
Db 27 GDKSSVVNNPT 37

RESULT 21

E85728

30S ribosomal subunit protein S22 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)

C;Species: Escherichia coli

C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 17-May-2002

C;Accession: E85728

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, G.F.; Evans, P.S.; Gregor, J.; Kirkpatrick, H.A.; Posfai, G.; Hackett, J.; Klink, S.; Boutin, A.; Shao, Y.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, J.; Anantharaman, T.S.; Lin, J.; Yen, G.; Schwartz, D.C.; Welch, R.A.; Blattner, F.R.  
Nature 409, 529-533, 2001

A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A;Reference number: A85480; MUID:21074935; PMID:11206551

A;Accession: E85728

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-45 <STO>

A;Cross-references: GB:AE005174; NID:g12515201; PIDN:AAG56289.1; GSPDB:GN00145; UWGP:Z2230

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: rpsV

C;Superfamily: Escherichia coli ribosomal protein S22

Query Match 18.4%; Score 39; DB 2; Length 45;  
Best Local Similarity 63.6%; Pred. No. 2e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 17 GQKSRVIENPT 27  
| || |: |||  
Db 27 GDKSSVVNNPT 37

RESULT 22

T06654

hypothetical protein T6G15.70 - Arabidopsis thaliana

C;Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 22-Oct-1999  
 C;Accession: T06654  
 R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.;  
 Mayer, K.F.X.; Schueller, C.  
 submitted to the Protein Sequence Database, April 1999  
 A;Reference number: Z15791  
 A;Accession: T06654  
 A;Molecule type: DNA  
 A;Residues: 1-62 <BEV>  
 A;Cross-references: EMBL:AL049656; GSPDB:GN00062; ATSP:T6G15.70  
 A;Experimental source: cultivar Columbia; BAC clone T6G15  
 C;Genetics:  
 A;Gene: ATSP:T6G15.70  
 A;Map position: 4

Query Match 18.4%; Score 39; DB 2; Length 62;  
 Best Local Similarity 29.7%; Pred. No. 2.9e+02;  
 Matches 11; Conservative 7; Mismatches 19; Indels 0; Gaps 0;

Qy 2 MRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGL 38  
 || : : | || : | : : || : : | ||  
 Db 1 MRPMQLDMLSEMDDAGSSMAMDVDDLEAMEILNEGGL 37

# RESULT 23

AD3532

hypothetical protein BMEII0182 [imported] - *Brucella melitensis* (strain 16M)

C;Species: *Brucella melitensis*

C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002

C;Accession: AD3532

R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.;  
 Ivanova, N.; Anderson, I.; Bhattacharyya, A.; Lykidis, A.; Reznik, G.;  
 Jablonski, L.; Larsen, N.; D'Souza, M.; Bernal, A.; Mazur, M.; Goltsman, E.;  
 Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letesson, J.J.; Haselkorn,  
 R.; Kyrpides, N.; Overbeek, R.

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A;Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*.

A;Reference number: AD3252; PMID:11756688

A;Accession: AD3532

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-72 <KUR>

A;Cross-references: GB:AE008918; PIDN:AAL53423.1; PID:g17984319; GSPDB:GN00191

A;Experimental source: strain 16M

C;Genetics:

A;Gene: BMEII0182

A;Map position: II

Query Match 18.4%; Score 39; DB 2; Length 72;  
 Best Local Similarity 41.7%; Pred. No. 3.5e+02;  
 Matches 10; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Qy 17 GQKSRVIENPTEALSVAVEEGLAW 40  
 | | | | | : : || || |  
 Db 4 GHLSYVRRNLVESRRLMVEIGLKW 27

RESULT 24

B83269

hypothetical protein PA3009 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C;Species: *Pseudomonas aeruginosa*

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C;Accession: B83269

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Brinkman, F.S.L.; Hufnagle, W.O.; Kowalik, D.J.; Lagrou, M.; Garber, R.L.; Goltry, L.; Tolentino, E.; Westbrook-Wadman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, R.M.; Smith, K.A.; Spencer, D.H.; Wong, G.K.S.; Wu, Z.; Paulsen, I.T.; Reizer, J.; Saier, M.H.; Hancock, R.E.W.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: B83269

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-77 <STO>

A;Cross-references: GB:AE004726; GB:AE004091; NID:g9949108; PIDN:AAG06397.1;

GSPDB:GN00131; PASP:PA3009

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA3009

Query Match	18.4%;	Score 39;	DB 2;	Length 77;
Best Local Similarity	31.8%;	Pred. No. 3.8e+02;		
Matches	7;	Conservative	6;	Mismatches 9; Indels 0; Gaps 0;

Qy	3	RSISENSLVAMDFSGQKSRVIE	24
		:   :   ::    ::	
Db	20	RADDEEALVTLEFSGDAKNFLQ	41

RESULT 25

T35253

small hypothetical protein SC5F2A.11 - *Streptomyces coelicolor*

C;Species: *Streptomyces coelicolor*

C;Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999

C;Accession: T35253

R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, April 1999

A;Reference number: Z21573

A;Accession: T35253

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-48 <OLI>

A;Cross-references: EMBL:AL049587; PIDN:CAB40678.1; GSPDB:GN00070;

SCOEDB:SC5F2A.11

A;Experimental source: strain A3(2)

C;Genetics:

A;Gene: SCOEDB:SC5F2A.11

Query Match 18.2%; Score 38.5; DB 2; Length 48;  
Best Local Similarity 36.7%; Pred. No. 2.5e+02;  
Matches 11; Conservative 4; Mismatches 14; Indels 1; Gaps 1;

Qy 12 AMDFSGQKSRVIENPTEA-LSVAVEEGLAW 40  
| : : | : | | : | | | |  
Db 6 APKYPARSGRPVERSVMVAGLVAVGAGLAW 35

Search completed: July 8, 2004, 08:20:45  
Job time : 13.1417 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2004, 08:20:54 ; Search time 27.7638 Seconds  
(without alignments)  
483.093 Million cell updates/sec

Title: US-09-936-697-5  
Perfect score: 212  
Sequence: 1 PMRSISENSLVAMDFSGQKS.....ENPTEALSVAVEEGLAWRKK 43

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1279676 seqs, 311918243 residues

Total number of hits satisfying chosen parameters: 487241

Minimum DB seq length: 0  
Maximum DB seq length: 85

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
------------	-------	--------------------	----	-------------

1	47.5	22.4	73	14	US-10-106-698-6971	Sequence 6971, Ap
2	47	22.2	68	16	US-10-437-963-202744	Sequence 202744,
3	45	21.2	84	9	US-09-864-761-47521	Sequence 47521, A
4	44.5	21.0	74	11	US-09-764-875-682	Sequence 682, App
5	44.5	21.0	74	11	US-09-764-875-998	Sequence 998, App
6	44	20.8	47	10	US-09-764-891-3024	Sequence 3024, Ap
7	43.5	20.5	36	12	US-10-424-599-205236	Sequence 205236,
8	43.5	20.5	55	9	US-09-879-957-111	Sequence 111, App
9	43.5	20.5	58	16	US-10-437-963-125413	Sequence 125413,
10	43.5	20.5	75	12	US-10-424-599-238531	Sequence 238531,
11	43	20.3	54	12	US-10-424-599-282889	Sequence 282889,
12	43	20.3	72	16	US-10-437-963-199279	Sequence 199279,
13	43	20.3	78	12	US-10-424-599-145859	Sequence 145859,
14	42.5	20.0	65	12	US-10-424-599-258371	Sequence 258371,
15	42.5	20.0	69	12	US-10-424-599-177050	Sequence 177050,
16	42.5	20.0	78	16	US-10-437-963-102594	Sequence 102594,
17	42	19.8	50	12	US-10-424-599-257895	Sequence 257895,
18	42	19.8	63	9	US-09-864-761-34262	Sequence 34262, A
19	42	19.8	63	12	US-10-424-599-270657	Sequence 270657,
20	42	19.8	70	12	US-10-282-122A-63145	Sequence 63145, A
21	42	19.8	70	14	US-10-268-518-5	Sequence 5, Appli
22	42	19.8	72	9	US-09-738-626-6764	Sequence 6764, Ap
23	42	19.8	76	12	US-10-282-122A-65522	Sequence 65522, A
24	42	19.8	78	12	US-10-424-599-145310	Sequence 145310,
25	42	19.8	80	12	US-10-424-599-156252	Sequence 156252,
26	42	19.8	81	15	US-10-289-762-1167	Sequence 1167, Ap
27	41.5	19.6	51	10	US-09-764-891-5151	Sequence 5151, Ap
28	41.5	19.6	64	12	US-10-424-599-243499	Sequence 243499,
29	41.5	19.6	67	16	US-10-437-963-116581	Sequence 116581,
30	41	19.3	53	12	US-10-424-599-271608	Sequence 271608,
31	41	19.3	55	16	US-10-437-963-162380	Sequence 162380,
32	41	19.3	67	12	US-10-424-599-143581	Sequence 143581,
33	41	19.3	69	12	US-10-424-599-220398	Sequence 220398,
34	41	19.3	73	12	US-10-424-599-215393	Sequence 215393,
35	41	19.3	76	16	US-10-437-963-131807	Sequence 131807,
36	41	19.3	79	16	US-10-408-765A-978	Sequence 978, App
37	40.5	19.1	50	16	US-10-437-963-125673	Sequence 125673,
38	40.5	19.1	51	14	US-10-097-111-420	Sequence 420, App
39	40.5	19.1	66	16	US-10-437-963-131467	Sequence 131467,
40	40.5	19.1	79	12	US-10-282-122A-49724	Sequence 49724, A
41	40	18.9	42	12	US-10-424-599-171821	Sequence 171821,
42	40	18.9	49	9	US-09-864-761-41352	Sequence 41352, A
43	40	18.9	63	10	US-09-983-802-611	Sequence 611, App
44	40	18.9	63	12	US-09-973-278-598	Sequence 598, App
45	40	18.9	63	12	US-09-984-490-611	Sequence 611, App
46	40	18.9	67	12	US-10-424-599-179446	Sequence 179446,
47	40	18.9	70	9	US-09-895-913A-280	Sequence 280, App
48	40	18.9	71	10	US-09-764-872-351	Sequence 351, App
49	40	18.9	76	12	US-10-424-599-214275	Sequence 214275,
50	40	18.9	78	12	US-10-424-599-207399	Sequence 207399,
51	40	18.9	78	12	US-10-425-114-48139	Sequence 48139, A
52	40	18.9	83	12	US-10-424-599-151847	Sequence 151847,
53	40	18.9	84	12	US-10-424-599-232889	Sequence 232889,
54	39.5	18.6	60	16	US-10-437-963-122272	Sequence 122272,
55	39.5	18.6	64	12	US-10-424-599-222312	Sequence 222312,
56	39.5	18.6	67	14	US-10-156-761-12354	Sequence 12354, A

57	39.5	18.6	68	9	US-09-864-761-47104	Sequence 47104, A
58	39.5	18.6	68	12	US-10-424-599-250542	Sequence 250542,
59	39.5	18.6	74	12	US-10-424-599-206895	Sequence 206895,
60	39.5	18.6	78	9	US-09-764-887-218	Sequence 218, App
61	39.5	18.6	78	14	US-10-073-961-218	Sequence 218, App
62	39	18.4	47	9	US-09-864-761-43815	Sequence 43815, A
63	39	18.4	52	12	US-10-424-599-158011	Sequence 158011,
64	39	18.4	59	12	US-10-424-599-249896	Sequence 249896,
65	39	18.4	60	9	US-09-864-761-45291	Sequence 45291, A
66	39	18.4	65	12	US-10-424-599-203529	Sequence 203529,
67	39	18.4	76	11	US-09-864-408A-6504	Sequence 6504, Ap
68	39	18.4	78	12	US-10-424-599-269854	Sequence 269854,
69	39	18.4	80	16	US-10-437-963-178166	Sequence 178166,
70	39	18.4	83	16	US-10-437-963-184263	Sequence 184263,
71	39	18.4	85	14	US-10-156-761-14857	Sequence 14857, A
72	38.5	18.2	30	10	US-09-829-382-5	Sequence 5, Appli
73	38.5	18.2	43	16	US-10-374-466-66	Sequence 66, Appl
74	38.5	18.2	51	14	US-10-327-514-21	Sequence 21, Appl
75	38.5	18.2	54	12	US-10-424-599-148148	Sequence 148148,
76	38.5	18.2	59	15	US-10-360-101-296	Sequence 296, App
77	38.5	18.2	63	12	US-10-424-599-178408	Sequence 178408,
78	38.5	18.2	66	9	US-09-864-761-36073	Sequence 36073, A
79	38.5	18.2	66	16	US-10-437-963-134503	Sequence 134503,
80	38.5	18.2	78	16	US-10-475-970-10	Sequence 10, Appl
81	38.5	18.2	79	16	US-10-437-963-168729	Sequence 168729,
82	38.5	18.2	85	9	US-09-522-334-28	Sequence 28, Appl
83	38.5	18.2	85	12	US-10-424-599-219916	Sequence 219916,
84	38	17.9	37	12	US-10-424-599-144034	Sequence 144034,
85	38	17.9	43	16	US-10-437-963-143842	Sequence 143842,
86	38	17.9	62	9	US-09-939-980-301	Sequence 301, App
87	38	17.9	63	12	US-10-424-599-189996	Sequence 189996,
88	38	17.9	68	9	US-09-925-300-1560	Sequence 1560, Ap
89	38	17.9	70	12	US-10-424-599-276208	Sequence 276208,
90	38	17.9	71	11	US-09-864-408A-604	Sequence 604, App
91	38	17.9	71	12	US-10-424-599-178736	Sequence 178736,
92	38	17.9	76	12	US-10-425-114-48440	Sequence 48440, A
93	38	17.9	77	12	US-10-424-599-221871	Sequence 221871,
94	38	17.9	79	11	US-09-864-408A-2680	Sequence 2680, Ap
95	38	17.9	79	12	US-10-424-599-240725	Sequence 240725,
96	38	17.9	80	12	US-10-424-599-165951	Sequence 165951,
97	38	17.9	80	12	US-10-424-599-226736	Sequence 226736,
98	38	17.9	80	12	US-10-424-599-268856	Sequence 268856,
99	38	17.9	83	12	US-10-424-599-151654	Sequence 151654,
100	38	17.9	84	12	US-10-424-599-168383	Sequence 168383,

#### ALIGNMENTS

##### RESULT 1

US-10-106-698-6971

; Sequence 6971, Application US/10106698

; Publication No. US20030109690A1

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides



; FILE REFERENCE: PA005P1  
 ; CURRENT APPLICATION NUMBER: US/10/106,698  
 ; CURRENT FILING DATE: 2002-03-27  
 ; PRIOR APPLICATION NUMBER: PCT/US00/26524  
 ; PRIOR FILING DATE: 2000-09-28  
 ; PRIOR APPLICATION NUMBER: US 60/157,137  
 ; PRIOR FILING DATE: 1999-09-29  
 ; PRIOR APPLICATION NUMBER: US 60/163,280  
 ; PRIOR FILING DATE: 1999-11-03  
 ; NUMBER OF SEQ ID NOS: 8564  
 ; SOFTWARE: PatentIn Ver. 3.0  
 ; SEQ ID NO 6971  
 ; LENGTH: 73  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-106-698-6971

Query Match 22.4%; Score 47.5; DB 14; Length 73;  
 Best Local Similarity 30.4%; Pred. No. 56;  
 Matches 14; Conservative 8; Mismatches 17; Indels 7; Gaps 2;

Qy 4 SISENSLVAMDFSGQKSRVIE-----NPTEAL--SVAVEEGLAWRK 42  
 :|||| :|: : : :| | |||| ||:| :  
 Db 11 TISENLFATTGYPGKMASQFQIHHLGHPQPILMGSVAVGSGLSWHR 56

## RESULT 2

US-10-437-963-202744

; Sequence 202744, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules  
 Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53221)B  
 ; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966  
 ; SEQ ID NO 202744

; LENGTH: 68  
 ; TYPE: PRT  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_97996C.1.pep  
 US-10-437-963-202744

Query Match 22.2%; Score 47; DB 16; Length 68;  
 Best Local Similarity 28.3%; Pred. No. 61;  
 Matches 13; Conservative 8; Mismatches 19; Indels 6; Gaps 2;

Qy 1 PMRSISENSL--VAMDFSGQKSRVIEN----PTEALSVAVEEGLAW 40  
 | | : | | : | | : | | : : : | |  
 Db 21 PPLSLSSHVLMPPVALSLDGHFSFRMITRVAPLPLELIGLVIRDGGGW 66

RESULT 3

US-09-864-761-47521

; Sequence 47521, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES  
 USEFUL FOR

; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

; FILE REFERENCE: Aeomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 47521  
; LENGTH: 84  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL158153.2  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 6.7  
; OTHER INFORMATION: EST\_HUMAN HIT: BF573955.1, EVALUE 1.60e-02  
; OTHER INFORMATION: SWISSPROT HIT: Q91641, EVALUE 3.00e-25  
US-09-864-761-47521

Query Match 21.2%; Score 45; DB 9; Length 84;  
Best Local Similarity 26.4%; Pred. No. 1.5e+02;  
Matches 14; Conservative 6; Mismatches 7; Indels 26; Gaps 2;

Qy 17 GQKSRVIENP-----TEALSVAV-----EEGLAWRKK 43  
|||:|:: | | ||: || | ||::  
Db 11 GQKARLLSRPLRGVSGKHCLTFFYHMYGGGTGLLSVYLKKEEDSEESLLWRRR 63

#### RESULT 4

US-09-764-875-682

; Sequence 682, Application US/09764875  
; Publication No. US20040018969A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PJZ02  
; CURRENT APPLICATION NUMBER: US/09/764,875  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1249  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 682  
; LENGTH: 74  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (8)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (55)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (61)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (68)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-875-682

Query Match 21.0%; Score 44.5; DB 11; Length 74;  
Best Local Similarity 23.6%; Pred. No. 1.5e+02;

Matches 13; Conservative 11; Mismatches 16; Indels 15; Gaps 2;

```
Qy      4 SISENSLVAMDFSGQKSRVIEN--PTEALSVA-----VEEGLAWRKK 43
      ||:|| |: |: |: :: | |: : : ||| |
Db      20 SITENGLIPKDYRSLKTQYLQSYGPEHLLTFSNLRXAGLLTXQAPGDNXTAWRVK 74
```

RESULT 5

US-09-764-875-998

```
; Sequence 998, Application US/09764875
; Publication No. US20040018969A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJZ02
; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 998
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (55)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (61)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (68)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-875-998
```

```
Query Match          21.0%; Score 44.5; DB 11; Length 74;
Best Local Similarity 23.6%; Pred. No. 1.5e+02;
Matches 13; Conservative 11; Mismatches 16; Indels 15; Gaps 2;
```

```
Qy      4 SISENSLVAMDFSGQKSRVIEN--PTEALSVA-----VEEGLAWRKK 43
      ||:|| |: |: |: :: | |: : : ||| |
Db      20 SITENGLIPKDYRSLKTQYLQSYGPEHLLTFSNLRXAGLLTXQAPGDNXTAWRVK 74
```

RESULT 6

US-09-764-891-3024

```
; Sequence 3024, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
```

; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3024  
; LENGTH: 47  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-891-3024

Query Match 20.8%; Score 44; DB 10; Length 47;  
Best Local Similarity 44.4%; Pred. No. 1e+02;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 20 SRVIENPTEALSVAVEEG 37  
|||:: || :|::| |  
Db 7 SRVLKGPTNIVSLSVNSG 24

RESULT 7

US-10-424-599-205236  
; Sequence 205236, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 205236  
; LENGTH: 36  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_27357C.1.pep  
US-10-424-599-205236

Query Match 20.5%; Score 43.5; DB 12; Length 36;  
Best Local Similarity 38.5%; Pred. No. 88;  
Matches 15; Conservative 7; Mismatches 10; Indels 7; Gaps 2;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLA 39  
|: ||| || | | : ||::|||:| | ::  
Db 2 PLSVISE--LVIRDSVQQ-----QLPTQSLSVSVSERMS 33

RESULT 8

US-09-879-957-111  
; Sequence 111, Application US/09879957  
; Patent No. US20020034755A1  
; GENERAL INFORMATION:

```

; APPLICANT: SPARKS, Andrew B.
;           HOFFMAN, No. US20020034755A1h
;           KAY, Brian K.
;           FOWLKES, Dana M.
;           McCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
;                   DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
;                   USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
;           ADDRESSEE: Pennie & Edmonds LLP
;           STREET: 1155 Avenue of the Americas
;           CITY: New York
;           STATE: New York
;           COUNTRY: USA
;           ZIP: 10036-2711
; COMPUTER READABLE FORM:
;           MEDIUM TYPE: Floppy disk
;           COMPUTER: IBM PC compatible
;           OPERATING SYSTEM: PC-DOS/MS-DOS
;           SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/09/879,957
;           FILING DATE: 13-Jun-2001
;           CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: US 08/630,915
;           FILING DATE: 03-APR-1996
; ATTORNEY/AGENT INFORMATION:
;           NAME: Misrock, S. Leslie
;           REGISTRATION NUMBER: 18,872
;           REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
;           TELEPHONE: (212) 790-9090
;           TELEFAX: (212) 869-8864/9741
;           TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 111:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 55 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: <Unknown>
;     TOPOLOGY: unknown
;   MOLECULE TYPE: peptide
;   SEQUENCE DESCRIPTION: SEQ ID NO: 111:
US-09-879-957-111

```

```

Query Match          20.5%;  Score 43.5;  DB 9;  Length 55;
Best Local Similarity 41.4%;  Pred. No. 1.5e+02;
Matches 12;  Conservative 8;  Mismatches 4;  Indels 5;  Gaps 2;

```

```

Qy      4 SISENSLVAMDFS-GQKSRVIENPTEALS 31
      :::: ||||: || ||::| | | | :
Db      23 TVNKGSLVALGFSDGQEAR----PEEILN 47

```

```

RESULT 9
US-10-437-963-125413

```



; LOCATION: (1)..(75)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_57419C.1.pep  
US-10-424-599-238531

Query Match 20.5%; Score 43.5; DB 12; Length 75;  
Best Local Similarity 34.6%; Pred. No. 2.2e+02;  
Matches 9; Conservative 6; Mismatches 4; Indels 7; Gaps 1;

Qy 16 SGQKSRVIENPTEALSVAVEEGLAWR 41  
||::||::|| | | ||  
Db 19 SGKQNRLLDPARACS-----TWR 37

RESULT 11

US-10-424-599-282889

; Sequence 282889, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated  
With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 282889  
; LENGTH: 54  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_97471C.1.pep  
US-10-424-599-282889

Query Match 20.3%; Score 43; DB 12; Length 54;  
Best Local Similarity 56.2%; Pred. No. 1.7e+02;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 17 GQKSRVIENPTEALSV 32  
||::| | ||:| |  
Db 1 GQRARKIFRPTKALGV 16

RESULT 12

US-10-437-963-199279

; Sequence 199279, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei



```
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 199279
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_9485C.1.pep
US-10-437-963-199279
```

```
Query Match          20.3%; Score 43; DB 16; Length 72;
Best Local Similarity 23.8%; Pred. No. 2.5e+02;
Matches 10; Conservative 9; Mismatches 13; Indels 10; Gaps 1;
```

```
Qy      1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRK 42
          | | : | |           | : | | : | : : : | : |
Db      37 PCRHVGERIL-----DVLVLPDESASLMIHDAVSWQK 68
```

```
RESULT 13
US-10-424-599-145859
; Sequence 145859, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 145859
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_10272C.1.pep
US-10-424-599-145859
```

```
Query Match          20.3%; Score 43; DB 12; Length 78;
Best Local Similarity 30.2%; Pred. No. 2.7e+02;
Matches 13; Conservative 10; Mismatches 14; Indels 6; Gaps 2;
```

```
Qy      1 PMRSISENSLVAMDFS---GQKSRVIENP---TEALSVAVEEG 37
```

Db                   | : |||| : : ||   | : : :|:   | :| : | :|  
34 PWQDISENVSLLLRFSYGLGETAYIIKTGLEITNSLQLIVRDG 76

RESULT 14

US-10-424-599-258371  
; Sequence 258371, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated  
With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 258371  
; LENGTH: 65  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_75333C.1.pep  
US-10-424-599-258371

Query Match                   20.0%; Score 42.5; DB 12; Length 65;  
Best Local Similarity   36.8%; Pred. No. 2.5e+02;  
Matches   14; Conservative   7; Mismatches   14; Indels   3; Gaps   2;

Qy           8 NSLIVAMDFSGQKSR--VIENPTEALSVAVEEGLAWRKK 43  
             ||           :|:|:| :|   ||   |: : :| | |||  
Db           26 NSPSVTTLNGRKTRSHLISEPTAHPSMLLQPGFA-RKK 62

RESULT 15

US-10-424-599-177050  
; Sequence 177050, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated  
With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 177050  
; LENGTH: 69  
; TYPE: PRT  
; ORGANISM: Glycine max

; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_130894C.1.pep  
US-10-424-599-177050

Query Match 20.0%; Score 42.5; DB 12; Length 69;  
Best Local Similarity 24.4%; Pred. No. 2.7e+02;  
Matches 10; Conservative 7; Mismatches 19; Indels 5; Gaps 1;

Qy 7 ENSLVAMDFSGQKSRVIENPTEALSVA-----VEEGLAWRK 42  
|| :| || :: | : ::| |||  
Db 5 ENENDGHSYSSAGSRTVKEPRVVQTTSEIDILDDGYRWRK 45

RESULT 16

US-10-437-963-102594  
; Sequence 102594, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules  
Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 102594  
; LENGTH: 78  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(78)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_100101C.1.pep  
US-10-437-963-102594

Query Match 20.0%; Score 42.5; DB 16; Length 78;  
Best Local Similarity 36.4%; Pred. No. 3.2e+02;  
Matches 16; Conservative 4; Mismatches 17; Indels 7; Gaps 3;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENP--TEALSVAVEEGLA-WR 41  
| | ||| | : | ||| | : :| :| | ||  
Db 22 PYRESXYNSLA---XGLQRRDWENPGVTQLISLAAHPPFASWR 61

RESULT 17

US-10-424-599-257895  
; Sequence 257895, Application US/10424599

; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated  
With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 257895  
; LENGTH: 50  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_74902C.1.pep  
US-10-424-599-257895

Query Match 19.8%; Score 42; DB 12; Length 50;  
Best Local Similarity 31.7%; Pred. No. 2.2e+02;  
Matches 13; Conservative 10; Mismatches 12; Indels 6; Gaps 2;

Qy 1 PMRSI-SENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAW 40  
|| |: | :|: : :|| | |:| :| ||  
Db 15 PMPSLKSHDSILNLERAGQHFAVTAQPSKA-----KEPDAW 50

# RESULT 18

US-09-864-761-34262  
; Sequence 34262, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES  
USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30

```

; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34262
; LENGTH: 63
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL031734.9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.87
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 74
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.82
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.79
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1
; OTHER INFORMATION: EST_HUMAN HIT: AI075970.1, EVALUE 7.00e-03
US-09-864-761-34262

```

```

Query Match          19.8%; Score 42; DB 9; Length 63;
Best Local Similarity 31.8%; Pred. No. 2.9e+02;
Matches 14; Conservative 7; Mismatches 11; Indels 12; Gaps 2;

```

```

Qy      6 SENSLVAMDFSGQKSRVI-----ENPTEALSVAVEEGLAWRK 42
      | : | : ||| : | : | | | : | | |
Db     13 SQVGLPILYFSGRRERLLLRPEVLAEIPREAFTVE-----AWVK 51

```

```

RESULT 19
US-10-424-599-270657
; Sequence 270657, Application US/10424599

```

; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated  
With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 270657  
; LENGTH: 63  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_86420C.1.pep  
US-10-424-599-270657

Query Match 19.8%; Score 42; DB 12; Length 63;  
Best Local Similarity 34.6%; Pred. No. 2.9e+02;  
Matches 9; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

Qy 15 FSGQKSRVIENPTEALSVAVEEGLAW 40  
| |:: | : | : || || |  
Db 20 FGRRRRRCYKGPSRRRLSPREEKEKW 45

# RESULT 20

US-10-282-122A-63145  
; Sequence 63145, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26

```

; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 63145
;   LENGTH: 70
;   TYPE: PRT
;   ORGANISM: Moraxella catarrhalis
US-10-282-122A-63145

```

```

Query Match          19.8%; Score 42; DB 12; Length 70;
Best Local Similarity 66.7%; Pred. No. 3.3e+02;
Matches      8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      28 EALSVAVEEGLA 39
        ||:||:|:|
Db      31 EAISVSLEDGLA 42

```

# RESULT 21

```

US-10-268-518-5
; Sequence 5, Application US/10268518
; Publication No. US20030100034A1
; GENERAL INFORMATION:
; APPLICANT: Hunter, John Joseph
; TITLE OF INVENTION: 9136, A HUMAN ALDEHYDE DEHYDROGENASE
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREFOR
; FILE REFERENCE: MPI01-234P1RM
; CURRENT APPLICATION NUMBER: US/10/268,518
; CURRENT FILING DATE: 2002-10-10
; PRIOR APPLICATION NUMBER: 60/329,899
; PRIOR FILING DATE: 2001-10-16
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 70
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Consensus sequence
US-10-268-518-5

```

```

Query Match          19.8%; Score 42; DB 14; Length 70;
Best Local Similarity 35.0%; Pred. No. 3.3e+02;

```

Matches 7; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Qy 23 IENPTEALSVAVEEGLAWRK 42  
:: || || : | ||:  
Db 37 VDKAVEAAQVAFQRGSPWRR 56

RESULT 22

US-09-738-626-6764

; Sequence 6764, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 6764  
; LENGTH: 72  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-6764

Query Match 19.8%; Score 42; DB 9; Length 72;  
Best Local Similarity 22.0%; Pred. No. 3.4e+02;  
Matches 11; Conservative 12; Mismatches 19; Indels 8; Gaps 1;

Qy 2 MRSISENSLVAMDFSGQKSRVI-----ENPTEALSVAVEEGLAWRKK 43  
| | || : : : : : : : : | : ||:||||:  
Db 1 MHFIKENLIFSASNALRAQLMLSILGSFAEFERSIIRERQAEGLAWRKR 50

RESULT 23

US-10-282-122A-65522

; Sequence 65522, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl



```

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 65522
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-282-122A-65522

```

```

Query Match          19.8%; Score 42; DB 12; Length 76;
Best Local Similarity 34.4%; Pred. No. 3.7e+02;
Matches 11; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

```

```

Qy      12 AMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43
        |:|: | :||:   |: |   || ||
Db       6 AVDYFGNESRL----ARAIGVKQPTVWAWNKK 33

```

```

RESULT 24
US-10-424-599-145310
; Sequence 145310, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J

```

```
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 145310
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(78)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_102234C.1.pep
US-10-424-599-145310
```

```
Query Match          19.8%; Score 42; DB 12; Length 78;
Best Local Similarity 61.5%; Pred. No. 3.8e+02;
Matches      8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy      16 SGQKSRVIENPTE 28
        | || |::| |||
Db      17 SQQKGRLVEXPTE 29
```

#### RESULT 25

```
US-10-424-599-156252
; Sequence 156252, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 156252
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112116C.1.pep
US-10-424-599-156252
```

```
Query Match          19.8%; Score 42; DB 12; Length 80;
Best Local Similarity 52.9%; Pred. No. 3.9e+02;
```

Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 25 NPTEALSVAVEEGLAWR 41  
| :|||:||| : ||  
Db 64 NMDALSLAVERIVDWR 80

Search completed: July 8, 2004, 08:31:40  
Job time : 28.7638 secs

OM protein - protein search, using sw model

Run on: July 8, 2004, 08:06:58 ; Search time 37.2441 Seconds  
(without alignments)  
364.280 Million cell updates/sec

Title: US-09-936-697-5  
Perfect score: 212  
Sequence: 1 PMRSISENSLVAMDFSGQKS.....ENPTEALSVAVEEGLAWRKK 43

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 123841

Minimum DB seq length: 0  
Maximum DB seq length: 85

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	Query	
No.	Score	Match Length DB ID Description
-----		

1	46.5	21.9	68	16	Q8E7B7	Q8e7b7 streptococc
2	46.5	21.9	83	16	Q8EIV8	Q8eiv8 shewanella
3	46	21.7	78	16	Q7VBG2	Q7vbg2 prochloroco
4	46	21.7	80	16	Q7V3F4	Q7v3f4 prochloroco
5	45.5	21.5	73	5	P91302	P91302 caenorhabdi
6	45.5	21.5	79	16	Q8EW35	Q8ew35 mycoplasma
7	45	21.2	59	9	Q855Q7	Q855q7 mycobacteri
8	45	21.2	77	17	O28902	O28902 archaeoglob
9	45	21.2	82	17	O27686	O27686 methanobact
10	44	20.8	60	17	Q8TTK0	Q8ttk0 methanosarc
11	44	20.8	79	2	Q9RCD4	Q9rcd4 xanthomonas
12	43.5	20.5	69	15	Q9WMQ6	Q9wmq6 human immun
13	43	20.3	76	16	Q835F1	Q835f1 enterococcu
14	42.5	20.0	80	3	Q9HGR8	Q9hgr8 choanephora
15	42	19.8	53	16	Q9PFG5	Q9pfg5 xylella fas
16	42	19.8	58	9	O80316	O80316 bacterioph
17	42	19.8	59	16	Q834Y7	Q834y7 enterococcu
18	42	19.8	60	17	Q8PWG8	Q8pwg8 methanosarc
19	42	19.8	72	16	Q8NLI4	Q8nli4 corynebacte
20	42	19.8	76	12	Q64947	Q64947 avian infec
21	42	19.8	76	12	Q64944	Q64944 avian infec
22	42	19.8	79	16	Q9Z8X5	Q9z8x5 chlamydia p
23	42	19.8	79	16	Q9JSH8	Q9jsh8 chlamydia p
24	42	19.8	81	16	Q9K247	Q9k247 chlamydia p
25	41.5	19.6	77	16	Q88KZ7	Q88kz7 pseudomonas
26	41	19.3	57	16	Q8EFI8	Q8efi8 shewanella
27	41	19.3	69	10	Q41693	Q41693 vigna radia
28	41	19.3	79	4	Q9NRP2	Q9nrp2 homo sapien
29	41	19.3	85	2	Q9WWG1	Q9wwg1 xanthomonas
30	41	19.3	85	16	Q8R9D7	Q8r9d7 thermoanaer
31	40.5	19.1	35	15	Q9W8L8	Q9w8l8 human immun
32	40.5	19.1	55	16	Q81UE2	Q8lue2 bacillus an
33	40.5	19.1	68	15	Q74620	Q74620 human immun
34	40.5	19.1	68	15	Q74630	Q74630 human immun
35	40.5	19.1	69	15	Q9WMQ5	Q9wmq5 human immun
36	40.5	19.1	69	15	Q9WMR4	Q9wmr4 human immun
37	40.5	19.1	74	10	Q8L8P5	Q8l8p5 arabidopsis
38	40.5	19.1	78	16	Q8F9J1	Q8f9j1 leptospira
39	40.5	19.1	79	2	Q936T5	Q936t5 pseudomonas
40	40	18.9	60	16	Q7VAL8	Q7val8 prochloroco
41	40	18.9	67	16	Q97IS7	Q97is7 clostridium
42	40	18.9	70	16	P71998	P71998 mycobacteri
43	40	18.9	70	16	Q7TZN6	Q7tzn6 mycobacteri
44	40	18.9	73	3	Q02288	Q02288 coccidioide
45	40	18.9	73	6	Q8MJD6	Q8mjd6 sus scrofa
46	40	18.9	73	13	Q8JHU0	Q8jhu0 gallus gall
47	40	18.9	80	16	Q8E852	Q8e852 shewanella
48	40	18.9	83	3	Q7Z879	Q7z879 talaromyces
49	40	18.9	84	12	Q9WKH1	Q9wkh1 encephalomy
50	39.5	18.6	35	15	Q9YM52	Q9ym52 human immun
51	39.5	18.6	56	2	Q9KK61	Q9kk61 mycobacteri
52	39.5	18.6	58	5	Q27193	Q27193 tetrahymena
53	39.5	18.6	67	16	Q92EM1	Q92em1 listeria in
54	39.5	18.6	67	16	Q9L0T9	Q9l0t9 streptomyce
55	39.5	18.6	67	16	Q82DZ6	Q82dz6 streptomyce
56	39.5	18.6	68	15	Q69653	Q69653 human immun
57	39.5	18.6	69	15	Q9WMR3	Q9wmr3 human immun

58	39.5	18.6	73	4	Q9BZL1	Q9bzl1 homo sapien
59	39.5	18.6	73	11	Q9EPV8	Q9epv8 mus musculu
60	39.5	18.6	73	13	Q7SXF2	Q7sxf2 brachydanio
61	39.5	18.6	77	16	Q8YXU8	Q8yxu8 anabaena sp
62	39.5	18.6	80	15	Q9QST4	Q9qst4 human immun
63	39.5	18.6	82	10	Q9LNN9	Q9lnn9 arabidopsis
64	39	18.4	41	6	O18852	O18852 macaca radi
65	39	18.4	41	16	Q8FCF2	Q8fcf2 escherichia
66	39	18.4	52	4	Q96GJ3	Q96gj3 homo sapien
67	39	18.4	54	10	Q8VY75	Q8vy75 arabidopsis
68	39	18.4	62	10	Q9T0H2	Q9t0h2 arabidopsis
69	39	18.4	67	4	Q9H1L3	Q9h1l3 homo sapien
70	39	18.4	67	16	Q81CY2	Q81cy2 bacillus ce
71	39	18.4	68	6	P79120	P79120 bos taurus
72	39	18.4	72	16	Q8YDJ3	Q8ydj3 brucella me
73	39	18.4	73	12	Q9YPL4	Q9ypl4 encephalomy
74	39	18.4	75	5	Q8IRX0	Q8irx0 drosophila
75	39	18.4	76	6	Q7YQJ3	Q7yqj3 bos taurus
76	39	18.4	76	12	Q64948	Q64948 avian infec
77	39	18.4	76	16	Q836W5	Q836w5 enterococcu
78	39	18.4	77	16	Q9HZJ7	Q9hzj7 pseudomonas
79	39	18.4	77	16	Q9K4K2	Q9k4k2 streptomyce
80	39	18.4	79	10	Q7XTI0	Q7xti0 oryza sativ
81	39	18.4	80	12	O91903	O91903 rupestris s
82	39	18.4	80	12	O91726	O91726 grapevine r
83	39	18.4	83	4	Q86X71	Q86x71 homo sapien
84	39	18.4	85	16	Q825W5	Q825w5 streptomyce
85	38.5	18.2	30	4	Q9HBG2	Q9hbg2 homo sapien
86	38.5	18.2	48	16	Q9X7N4	Q9x7n4 streptomyce
87	38.5	18.2	55	16	Q81H66	Q81h66 bacillus ce
88	38.5	18.2	67	16	Q8Y9V4	Q8y9v4 listeria mo
89	38.5	18.2	67	16	Q8Y4M8	Q8y4m8 listeria mo
90	38.5	18.2	68	15	Q74093	Q74093 human immun
91	38.5	18.2	68	15	Q74645	Q74645 human immun
92	38.5	18.2	69	2	Q51902	Q51902 proteus mir
93	38.5	18.2	69	11	Q9CWL5	Q9cwl5 mus musculu
94	38.5	18.2	69	15	Q9WMT2	Q9wmt2 human immun
95	38.5	18.2	74	16	Q8P1Q2	Q8plq2 streptococc
96	38.5	18.2	80	14	Q99IV0	Q99iv0 uncultured
97	38.5	18.2	80	16	Q82UB4	Q82ub4 nitrosomona
98	38.5	18.2	81	12	Q7TE81	Q7te81 dolichos ye
99	38.5	18.2	81	15	Q90QQ6	Q90qq6 human immun
100	38.5	18.2	83	10	Q8GZL0	Q8gzl0 arabidopsis

#### ALIGNMENTS

##### RESULT 1

Q8E7B7

ID Q8E7B7 PRELIMINARY; PRT; 68 AA.

AC Q8E7B7;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Hypothetical protein.

GN GBS0238.

OS Streptococcus agalactiae (serotype III).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=216495;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NEM316 / Serotype III;  
 RX MEDLINE=22242508; PubMed=12354221;  
 RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,  
 RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,  
 RA Kunst F.;  
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing  
 RT invasive neonatal disease.";  
 RL Mol. Microbiol. 45:1499-1513(2002).  
 DR EMBL; AL766844; CAD45883.1; -.  
 DR SagaList; gbs0238; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 68 AA; 7450 MW; 33108A42C112BF80 CRC64;

Query Match 21.9%; Score 46.5; DB 16; Length 68;  
 Best Local Similarity 39.3%; Pred. No. 1.5e+02;  
 Matches 11; Conservative 6; Mismatches 10; Indels 1; Gaps 1;

Qy 4 SISENSLVAMDFS-GQKSRVIENPTEAL 30  
 :|::| |:|: || | |:| | |  
 Db 4 TINKNDLIALGFSEGTSKRIIRQKELL 31

## RESULT 2

### Q8EIV8

ID Q8EIV8 PRELIMINARY; PRT; 83 AA.  
 AC Q8EIV8;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Conserved hypothetical protein.  
 GN SO0721.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadaceae; Shewanella.  
 OX NCBI\_TaxID=70863;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MR-1;  
 RX MEDLINE=22297686; PubMed=12368813;  
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,  
 RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
 RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,  
 RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,  
 RA Feldblyum T.V., Smith H.O., Venter J.C., Nealson K.H., Fraser C.M.;  
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 RT Shewanella oneidensis.";  
 RL Nat. Biotechnol. 20:1118-1123(2002).  
 DR EMBL; AE015517; AAN53799.1; -.

DR TIGR; SO0721; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 83 AA; 9075 MW; AC5D08F38ACB345C CRC64;

Query Match 21.9%; Score 46.5; DB 16; Length 83;  
Best Local Similarity 25.5%; Pred. No. 1.8e+02;  
Matches 13; Conservative 11; Mismatches 16; Indels 11; Gaps 1;

Qy 3 RSISENSLVAMDFSGQ-----KSRVIENPTEALSVAVEEGLAWRK 42  
:::| |::| | : |::| : | | : |  
Db 23 QALTDNPLMAMGIIGQLGIPPEKLQQLMALVMQNPALIKEAVLELGLDFAK 73

#### RESULT 3

##### Q7VBG2

ID Q7VBG2 PRELIMINARY; PRT; 78 AA.  
AC Q7VBG2;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Predicted protein.  
GN PRO1133.  
OS Prochlorococcus marinus.  
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;  
OC Prochlorococcus.  
OX NCBI\_TaxID=1219;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SARG / CCMP 1375 / SS120;  
RX MEDLINE=22810154; PubMed=12917486;  
RA Dufresne A., Salanoubat M., Partensky F., Artiguenave F., Axmann I.M.,  
RA Barbe V., Duprat S., Galperin M.Y., Koonin E.V., Le Gall F.,  
RA Makarova K.S., Ostrowski M., Oztas S., Robert C., Rogozin I.B.,  
RA Scanlan D.J., Tandeau de Marsac N., Weissenbach J., Wincker P.,  
RA Wolf Y.I., Hess W.R.;  
RT "Genome sequence of the cyanobacterium Prochlorococcus marinus SS120,  
RT a nearly minimal oxyphototrophic genome."  
RL Proc. Natl. Acad. Sci. U.S.A. 100:10020-10025(2003).  
DR EMBL; AE017164; AAQ00178.1; -.  
KW Complete proteome.  
SQ SEQUENCE 78 AA; 8555 MW; 338B0D6AE8B40155 CRC64;

Query Match 21.7%; Score 46; DB 16; Length 78;  
Best Local Similarity 32.4%; Pred. No. 2e+02;  
Matches 11; Conservative 5; Mismatches 16; Indels 2; Gaps 1;

Qy 10 LVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
|| || | |::| |::| : | |  
Db 19 LVGMD--GHPHPVLDTPYESVDAAIGAAKQWTSK 50

#### RESULT 4

##### Q7V3F4

ID Q7V3F4 PRELIMINARY; PRT; 80 AA.  
AC Q7V3F4;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)



DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN PMM0121.  
 OS Prochlorococcus marinus subsp. pastoris (strain CCMP 1378 / MED4).  
 OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;  
 OC Prochlorococcus.  
 OX NCBI\_TaxID=59919;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22825698; PubMed=12917642;  
 RA Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,  
 RA Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,  
 RA Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,  
 RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,  
 RA Webb E.A., Zinser E.R., Chisholm S.W.;  
 RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic  
 RT niche differentiation."  
 RL Nature 424:1042-1047(2003).  
 DR EMBL; BX572090; CAE18580.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 80 AA; 9218 MW; 19A642863632D7CA CRC64;  
  
 Query Match 21.7%; Score 46; DB 16; Length 80;  
 Best Local Similarity 26.8%; Pred. No. 2.1e+02;  
 Matches 15; Conservative 9; Mismatches 18; Indels 14; Gaps 3;  
  
 Qy 1 PMRSISENSLIVAMD---FSGQ-----KSRVIENPTEALSVAVEEG-LAWRK 42  
 | : : : ||| : | : | : | | || : | : ||:  
 Db 6 PKKPLKKGSLVFIDKSIYDGSVEALASDQDLPSYIFEGPGEILSIKEEYAQVRWRR 61

# RESULT 5

P91302  
 ID P91302 PRELIMINARY; PRT; 73 AA.  
 AC P91302;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE F46F11.4 protein.  
 GN F46F11.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium."  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Pauley A., Gattung S.;  
 RT "The sequence of C. elegans cosmid F46F11."

RL Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; U88173; AAK21382.1; -.  
 DR PIR; T25763; T25763.  
 DR WormPep; F46F11.4; CE10602.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00240; ubiquitin; 1.  
 DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
 SQ SEQUENCE 73 AA; 8738 MW; 61CA839BBA4006A4 CRC64;

Query Match 21.5%; Score 45.5; DB 5; Length 73;  
 Best Local Similarity 29.4%; Pred. No. 2.2e+02;  
 Matches 10; Conservative 7; Mismatches 12; Indels 5; Gaps 1;

Qy 14 DFSGQKSRVIENPTEALS-----VAVEEGLAWRK 42  
 | |:| |: ||:: : :| :| ||  
 Db 8 DRLGKKVRIKCNPSDTIGDLKKLIAAQTGTRWEK 41

#### RESULT 6

##### Q8EW35

ID Q8EW35 PRELIMINARY; PRT; 79 AA.  
 AC Q8EW35;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN MYPE3720.  
 OS Mycoplasma penetrans.  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 OX NCBI\_TaxID=28227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HF-2;  
 RX MEDLINE=22354719; PubMed=12466555;  
 RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,  
 RA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;  
 RT "The complete genomic sequence of Mycoplasma penetrans, an  
 RT intracellular bacterial pathogen in humans.";  
 RL Nucleic Acids Res. 30:5293-5300(2002).  
 DR EMBL; AP004171; BAC44161.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 79 AA; 9655 MW; 357C5690D747E091 CRC64;

Query Match 21.5%; Score 45.5; DB 16; Length 79;  
 Best Local Similarity 37.0%; Pred. No. 2.4e+02;  
 Matches 10; Conservative 7; Mismatches 7; Indels 3; Gaps 1;

Qy 10 LVAMDFSGQKSRVIENPTEALSVAVEE 36  
 || ||:| :| :|| ::| :|  
 Db 47 LVREDFNG---KVFKNPEHNITIIVDE 70

# RESULT 7

Q855Q7

ID Q855Q7 PRELIMINARY; PRT; 59 AA.  
AC Q855Q7;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Gp56.  
OS Mycobacteriophage Che9d.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
OX NCBI\_TaxID=205876;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22592660; PubMed=12705866;  
RA Pedulla M.L., Ford M.E., Houtz J.M., Karthikeyan T., Wadsworth C.,  
RA Lewis J.A., Jacobs-Sera D., Falbo J., Gross J., Pannunzio N.R.,  
RA Brucker W., Kumar V., Kandasamy J., Keenan L., Bardarov S.,  
RA Kriakov J., Lawrence J.G., Jacobs W.R. Jr., Hendrix R.W.,  
RA Hatfull G.F.;  
RT "Origins of highly mosaic mycobacteriophage genomes."  
RL Cell 113:171-182(2003).  
DR EMBL; AY129336; AAN07974.1; -.  
SQ SEQUENCE 59 AA; 6611 MW; 7D7AAFBD6743760 CRC64;

Query Match 21.2%; Score 45; DB 9; Length 59;  
Best Local Similarity 23.1%; Pred. No. 2e+02;  
Matches 9; Conservative 11; Mismatches 19; Indels 0; Gaps 0;

Qy 3 RSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
| : : : | : | : : || : | :  
Db 4 RLLYDKAAAQVLSTERRIDELRRAGVLIQVQDGREWK 42

# RESULT 8

O28902

ID O28902 PRELIMINARY; PRT; 77 AA.  
AC O28902;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hydrogenase expression/formation protein (HYPC).  
GN AF1369.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
OX NCBI\_TaxID=2234;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,  
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,

RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon *Archaeoglobus fulgidus*.";  
 RL Nature 390:364-370(1997).  
 DR EMBL; AE001009; AAB89878.1; -.  
 DR TIGR; AF1369; -.  
 DR InterPro; IPR001109; HupF\_HypC.  
 DR Pfam; PF01455; HupF\_HypC; 1.  
 DR PRINTS; PR00445; HUPFHYP\_C.  
 DR ProDom; PD003112; HupF\_HypC; 1.  
 DR TIGRFAMs; TIGR00074; hypC\_hupF; 1.  
 DR PROSITE; PS01097; HUPF\_HYP\_C; 1.  
 DR PIRSF; PIRSF005618; HupF\_HypC; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 77 AA; 8783 MW; 669179CCB544D027 CRC64;

Query Match 21.2%; Score 45; DB 17; Length 77;  
 Best Local Similarity 35.1%; Pred. No. 2.7e+02;  
 Matches 13; Conservative 5; Mismatches 15; Indels 4; Gaps 1;

Qy 10 LVAMDFSGQKSRV----IENPTEALSVAVEEGLAWRK 42  
 : :||| | | :||| | | |:| :|  
 Db 16 IAIVDFKGLKKEVRIDLLENPQIGDYVLVHVGMAIQK 52

# RESULT 9

O27686

ID O27686 PRELIMINARY; PRT; 82 AA.  
 AC O27686;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hydrogenase expression/formation protein HYP\_C.  
 GN MTH1649.  
 OS *Methanobacterium thermoautotrophicum*.  
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
 OC Methanobacteriaceae; Methanothermobacter.  
 OX NCBI\_TaxID=187420;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Delta H;  
 RX MEDLINE=98037514; PubMed=9371463;  
 RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,  
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
 RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
 RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,  
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;  
 RT "Complete genome sequence of *Methanobacterium thermoautotrophicum*  
 RT deltaH: functional analysis and comparative genomics.";  
 RL J. Bacteriol. 179:7135-7155(1997).  
 DR EMBL; AE000924; AAB86122.1; -.  
 DR InterPro; IPR001109; HupF\_HypC.

DR Pfam; PF01455; HupF\_HypC; 1.  
 DR PRINTS; PR00445; HUPFHYP C.  
 DR ProDom; PD003112; HupF\_HypC; 1.  
 DR TIGRFAMs; TIGR00074; hypC\_hupF; 1.  
 DR PIRSF; PIRSF005618; HupF\_HypC; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 82 AA; 9082 MW; B6E6AED010FBE62D CRC64;

Query Match 21.2%; Score 45; DB 17; Length 82;  
 Best Local Similarity 28.9%; Pred. No. 2.9e+02;  
 Matches 11; Conservative 9; Mismatches 14; Indels 4; Gaps 1;

QY 6 SENSLVAMDFSGQKSRV----IENPTEALSVAVEEGLA 39  
 ||::: :|| | : :| :: | | | |  
 Db 14 SEDNIATVDFGGVRQVKLDLVDVVEEGKYVLVHSGYA 51

# RESULT 10

## Q8TTK0

ID Q8TTK0 PRELIMINARY; PRT; 60 AA.  
 AC Q8TTK0;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Ferredoxin.  
 GN MA0431.  
 OS Methanosarcina acetivorans.  
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;  
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.  
 OX NCBI\_TaxID=2214;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;  
 RX MEDLINE=21929760; PubMed=11932238;  
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,  
 RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,  
 RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,  
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,  
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,  
 RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,  
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,  
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,  
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,  
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,  
 RA Metcalf W.W., Birren B.;  
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic  
 RT and physiological diversity.";  
 RL Genome Res. 12:532-542(2002).  
 DR EMBL; AE010703; AAM03878.1; -.  
 DR GO; GO:0005489; F:electron transporter activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR001450; 4Fe4S\_ferredoxin.  
 DR Pfam; PF00037; fer4; 2.  
 DR PROSITE; PS00198; 4FE4S\_FERREDOXIN; 2.  
 KW Complete proteome.  
 SQ SEQUENCE 60 AA; 6265 MW; 6D75EBDB4460C21F CRC64;

Query Match 20.8%; Score 44; DB 17; Length 60;  
Best Local Similarity 42.9%; Pred. No. 2.7e+02;  
Matches 12; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

Qy 12 AMDFSGQKSRVIENPTEALSVAVEEGLA 39  
| : || : | | | : || : : | : |||  
Db 7 ADECSGCGTCVDECPSEAITLDEEKGLA 34

RESULT 11

Q9RCD4

ID Q9RCD4 PRELIMINARY; PRT; 79 AA.  
AC Q9RCD4;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Xanthomonas campestris.  
OG Plasmid pKLH443.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xanthomonas.  
OX NCBI\_TaxID=339;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TAP44-3; TRANSPOSON=Tn5044;  
RX MEDLINE=99406912; PubMed=10476039;  
RA Minakhina S., Kholodii G., Mindlin S., Yurieva O., Nikiforov V.;  
RT "Tn5053 family transposons are res site hunters sensing plasmidal res  
RT sites occupied by cognate resolvases.";  
RL Mol. Microbiol. 33:1059-1068(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TAP44-3; TRANSPOSON=Tn5044;  
RA Kholodii G., Yurieva O., Mindlin S., Gorlenko Z., Rybochkin V.,  
RA Nikiforov V.;  
RT "Tn5044, a novel Tn3 family transposon coding for temperature  
RT sensitive mercury resistance.";  
RL Res. Microbiol. 151:1-12(2000).  
DR EMBL; Y17691; CAB65713.1; -.  
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 79 AA; 8626 MW; 1639B3E026E36706 CRC64;

Query Match 20.8%; Score 44; DB 2; Length 79;  
Best Local Similarity 33.3%; Pred. No. 3.7e+02;  
Matches 11; Conservative 11; Mismatches 9; Indels 2; Gaps 2;

QY 13 MDFSGQKSRVIE-NPTEA-LSVAVEEGLAWRKK 43  
:: | :::: : : : || | : ||: | || |  
Db 46 LELSAEQAKAVNAHLSEAELTDAVDEALAWASK 78

RESULT 12

Q9WMQ6

ID Q9WMQ6 PRELIMINARY; PRT; 69 AA.  
AC Q9WMQ6;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Reverse transcriptase (Fragment).  
 GN RT.  
 OS Human immunodeficiency virus.  
 OC Viruses; Retrovird viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=12721;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DRO22/M0;  
 RX MEDLINE=20146732; PubMed=10682151;  
 RA Masquelier B., Descamps D., Carriere I., Ferchal F., Collin G.,  
 RA Denayrolles M., Ruffault A., Chanzy B., Izopet J.,  
 RA Buffet-Janvresse C., Schmitt M.P., Race E., Fleury H.J.A.,  
 RA Aboulker J.P., Yeni P., Brun-Vezinet F.;  
 RT "Resensitization and dual HIV-1 resistance to zidovudine and  
 RT lamivudine in the Delta lamivudine roll-over study.";  
 RL Antivir. Ther. 4:69-77(1999).  
 DR EMBL; AJ239270; CAB51518.1; -.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006278; P:RNA dependent DNA replication; IEA.  
 DR InterPro; IPR000477; RVTse.  
 DR Pfam; PF00078; rvt; 1.  
 KW RNA-directed DNA polymerase; Transferase.  
 FT NON\_TER 1 1  
 FT NON\_TER 69 69  
 SQ SEQUENCE 69 AA; 8089 MW; 5BD8FF800A16A70C CRC64;

Query Match 20.5%; Score 43.5; DB 15; Length 69;  
 Best Local Similarity 33.3%; Pred. No. 3.7e+02;  
 Matches 11; Conservative 8; Mismatches 11; Indels 3; Gaps 2;

Qy 13 MDFSGQKSRV-IENP--TEALSVAVEEGLAWRK 42  
 :: |: |:: ||| | :: :|| |||  
 Db 15 LEKEGKISKIGPENPYNTPVFAIKKKEGTKWRK 47

# RESULT 13

Q835F1

ID Q835F1 PRELIMINARY; PRT; 76 AA.  
 AC Q835F1;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE VrlI protein, putative.  
 GN EF1426.  
 OS Enterococcus faecalis (Streptococcus faecalis).  
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.  
 OX NCBI\_TaxID=1351;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=v583 / ATCC 700802;  
 RX MEDLINE=22550857; PubMed=12663927;  
 RA Paulsen I.T., Banerjei L., Myers G.S.A., Nelson K.E., Seshadri R.,  
 RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,

RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,  
 RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,  
 RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,  
 RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;  
 RT "Role of mobile DNA in the evolution of vancomycin-resistant  
 RT Enterococcus faecalis.";  
 RL Science 299:2071-2074(2003).  
 DR EMBL; AE016951; AAO81217.1; -.  
 DR TIGR; EF1426; -.  
 KW Complete proteome.  
 SQ SEQUENCE 76 AA; 8880 MW; E2CFCF862B3C2795 CRC64;

Query Match 20.3%; Score 43; DB 16; Length 76;  
 Best Local Similarity 24.4%; Pred. No. 4.9e+02;  
 Matches 11; Conservative 11; Mismatches 13; Indels 10; Gaps 2;

Qy 9 SLVAMDFSGQKSRVIEN----PTEALSVAVEEGLA-----WRKK 43  
 :| :|| :| | : : | :| :|| : : : | | :  
 Db 2 ALEVIDFKSKKDRKVNNSKIPPLKAIEVAKRKNVSAATVTRWMKR 46

#### RESULT 14

##### Q9HGR8

ID Q9HGR8 PRELIMINARY; PRT; 80 AA.  
 AC Q9HGR8;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH)  
 DE (Fragment).  
 GN GPD.  
 OS Choanephora infundibulifera.  
 OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Choanephoraceae;  
 OC Choanephora.  
 OX NCBI\_TaxID=127959;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL 2560;  
 RA Tamas P.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL 2560;  
 RA Papp T., Vastag M., Acs K., Vagvolgyi C.;  
 RT "Phylogenetic relationships among Mucoraceae, Choanephoraceae and  
 RT Gilbertellaceae based on rDNA and glyceraldehyde-3-phosphate  
 RT dehydrogenase sequences.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 CC -!- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + PHOSPHATE +  
 CC NAD(+) = 3-PHOSPHO-D-GLYCEROYL PHOSPHATE + NADH.  
 CC -!- PATHWAY: SECOND PHASE OF GLYCOLYSIS; FIRST STEP.  
 CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE  
 CC DEHYDROGENASE FAMILY.  
 DR EMBL; AJ278315; CAC05662.1; -.  
 DR HSSP; P00354; 3GPD.



DR GO; GO:0004365; F:glyceraldehyde-3-phosphate dehydrogenase (p. . .; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0006096; P:glycolysis; IEA.  
 DR InterPro; IPR000173; GAP\_dhhydrogenase.  
 DR Pfam; PF02800; gpdh\_C; 1.  
 KW Glycolysis; NAD; Oxidoreductase.  
 FT NON\_TER 1 1  
 FT NON\_TER 80 80  
 SQ SEQUENCE 80 AA; 8504 MW; 4ECCBEAE035943D0 CRC64;

Query Match 20.0%; Score 42.5; DB 3; Length 80;  
 Best Local Similarity 42.9%; Pred. No. 6e+02;  
 Matches 9; Conservative 6; Mismatches 3; Indels 3; Gaps 1;

Qy 1 PMRSI---SENSLVAMDFSGQ 18  
 ||: | :||::|: || |:   
 Db 37 PMKGILGYTENAVVSTDFIGE 57

# RESULT 15

## Q9PFG5

ID Q9PFG5 PRELIMINARY; PRT; 53 AA.  
 AC Q9PFG5;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein Xf0694.  
 GN XF0694.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xylella.  
 OX NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9a5c;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,

RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*.";  
 RL Nature 406:151-159(2000).  
 DR EMBL; AE003912; AAF83504.1; -.  
 DR PIR; C82776; C82776.  
 DR InterPro; IPR000437; Prok\_lipoprot\_S.  
 DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 53 AA; 5958 MW; 4B14AF832900832B CRC64;

Query Match 19.8%; Score 42; DB 16; Length 53;  
 Best Local Similarity 66.7%; Pred. No. 4.4e+02;  
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 30 LSVAVEEGLAWR 41  
 | | | | | | |  
 Db 21 LGVGVERGYAWR 32

#### RESULT 16

O80316

ID O80316 PRELIMINARY; PRT; 58 AA.  
 AC O80316;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Orf52 (Fragment).  
 GN H.  
 OS Bacteriophage 186.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
 OC P2-like viruses.  
 OX NCBI\_TaxID=29252;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Xue Q.;  
 RT "Studies on the tail region of the temperate coliphage 186 genome.";  
 RL Thesis (1993), University of Adelaide.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98371265; PubMed=9705261;  
 RA Portelli R., Dodd I.B., Xue Q., Egan J.B.;  
 RT "The late-expressed region of the temperate coliphage 186 genome.";  
 RL Virology 248:117-130(1998).  
 DR EMBL; U32222; AAC34169.1; -.  
 FT NON\_TER 1 1  
 FT VARIANT 15 15 S -> \*.  
 FT VARIANT 51 51 Q -> \*.  
 SQ SEQUENCE 58 AA; 6491 MW; 1199113D8CDEB8E6 CRC64;

Query Match 19.8%; Score 42; DB 9; Length 58;  
 Best Local Similarity 38.9%; Pred. No. 4.9e+02;  
 Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 26 PTEALSVAVEEGLAWRKK 43  
 |::| |::: | : ||:|  
 Db 31 PSELYSLSLTELITWREK 48

RESULT 17

Q834Y7

ID Q834Y7 PRELIMINARY; PRT; 59 AA.  
 AC Q834Y7;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 GN EF1490.  
 OS Enterococcus faecalis (Streptococcus faecalis).  
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.  
 OX NCBI\_TaxID=1351;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=V583 / ATCC 700802;  
 RX MEDLINE=22550857; PubMed=12663927;  
 RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,  
 RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,  
 RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,  
 RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,  
 RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,  
 RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;  
 RT "Role of mobile DNA in the evolution of vancomycin-resistant  
 RT Enterococcus faecalis."  
 RL Science 299:2071-2074(2003).  
 DR EMBL; AE016951; AAO81281.1; -.  
 DR TIGR; EF1490; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 59 AA; 6993 MW; E40B1722F9E762F1 CRC64;

Query Match 19.8%; Score 42; DB 16; Length 59;  
 Best Local Similarity 34.1%; Pred. No. 4.9e+02;  
 Matches 15; Conservative 11; Mismatches 14; Indels 4; Gaps 3;

Qy 2 MRSISE--NSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 ::||| | | : : : : : | : | | : | | |  
 Db 10 LQSISEEPNSFI-TEETIKYIEQLEDDNESLQVAL-EGTIWSPK 51

RESULT 18

Q8PWG8

ID Q8PWG8 PRELIMINARY; PRT; 60 AA.  
 AC Q8PWG8;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Ferredoxin.  
 GN MM1619.  
 OS Methanosarcina mazei (Methanosarcina frisia).  
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;  
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.  
 OX NCBI\_TaxID=2209;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Goel / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;  
 RX MEDLINE=22120827; PubMed=12125824;  
 RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,  
 RA Martinez-Arias R., Henne A., Wiezer A., Baeumer S., Jacobi C.,  
 RA Brueggemann H., Lienard T., Christmann A., Boemecke M., Steckel S.,  
 RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,  
 RA Fritz H.-J., Gottschalk G.;  
 RT "The genome of Methanosarcina mazei: evidence for lateral gene  
 RT transfer between Bacteria and Archaea.";  
 RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).  
 DR EMBL; AE013395; AAM31315.1; -.  
 DR GO; GO:0005489; F:electron transporter activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR001450; 4Fe4S\_ferredoxin.  
 DR Pfam; PF00037; fer4; 2.  
 DR PROSITE; PS00198; 4FE4S\_FERREDOXIN; 2.  
 KW Complete proteome.  
 SQ SEQUENCE 60 AA; 6237 MW; 6D6F5BDE1435C21F CRC64;  
  
 Query Match 19.8%; Score 42; DB 17; Length 60;  
 Best Local Similarity 39.3%; Pred. No. 5e+02;  
 Matches 11; Conservative 8; Mismatches 9; Indels 0; Gaps 0;  
  
 Qy 12 AMDFSGQKSRVIENPTEALSVAVEEGLA 39  
 | : || : | | | : || :: | : | |  
 Db 7 ADECSGCGTCVDECPSEAITLDEEKGIA 34

# RESULT 19

Q8NLI4  
 ID Q8NLI4 PRELIMINARY; PRT; 72 AA.  
 AC Q8NLI4;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Site-specific recombinases, DNA invertase Pin homologs.  
 GN CGL2958.  
 OS Corynebacterium glutamicum (Brevibacterium flavum).  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 OX NCBI\_TaxID=1718;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;  
 RA Nakagawa S.;  
 RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AP005283; BAC00352.1; -.  
 DR GO; GO:0000150; F:recombinase activity; IEA.  
 DR GO; GO:0006310; P:DNA recombination; IEA.  
 DR InterPro; IPR006119; resolvase\_N.  
 DR Pfam; PF00239; resolvase; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 72 AA; 8042 MW; A4F4F84F57B17F07 CRC64;

Query Match 19.8%; Score 42; DB 16; Length 72;  
Best Local Similarity 22.0%; Pred. No. 6.2e+02;  
Matches 11; Conservative 12; Mismatches 19; Indels 8; Gaps 1;

Qy 2 MRSISENSLVAMDFSGQKSRVI-----ENPTEALSVAVEEGLAWRKK 43  
| | || : : : : : : : | : ||:||||:  
Db 1 MHFIKENLIFSAESNALRAQLMLSILGSFAEFERSIIRERQAEGIAWRKR 50

RESULT 20

Q64947

ID Q64947 PRELIMINARY; PRT; 76 AA.  
AC Q64947;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE Spike protein (Fragment).  
GN S1.  
OS Avian infectious bronchitis virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
OC Coronaviridae; Coronavirus.  
OX NCBI\_TaxID=11120;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A1955;  
RX MEDLINE=97049060; PubMed=8893790;  
RA Wang C.H., Tsai C.T.;  
RT "Genetic grouping for the isolates of avian infectious bronchitis  
virus in Taiwan."  
RL Arch. Virol. 141:1677-1688(1996).  
DR EMBL; U38681; AAB47439.1; -.  
DR InterPro; IPR002551; Corona\_S1.  
DR Pfam; PF01600; Corona\_S1; 1.  
FT NON\_TER 1 1  
FT NON\_TER 76 76  
SQ SEQUENCE 76 AA; 7903 MW; 271F114FD4078521 CRC64;

Query Match 19.8%; Score 42; DB 12; Length 76;  
Best Local Similarity 36.4%; Pred. No. 6.6e+02;  
Matches 8; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 21 RVIENTEALSVAVEEGLAWRK 42  
||: : |:| |:| :| :|  
Db 36 RVVNASSIAMSAFVGQGMQWSK 57

RESULT 21

Q64944

ID Q64944 PRELIMINARY; PRT; 76 AA.  
AC Q64944;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE Spike protein (Fragment).  
GN S1.  
OS Avian infectious bronchitis virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;

OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxID=11120;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A1960;  
 RX MEDLINE=97049060; PubMed=8893790;  
 RA Wang C.H., Tsai C.T.;  
 RT "Genetic grouping for the isolates of avian infectious bronchitis  
 RT virus in Taiwan.";  
 RL Arch. Virol. 141:1677-1688(1996).  
 DR EMBL; U38678; AAB47436.1; -.  
 DR InterPro; IPR002551; Corona\_S1.  
 DR Pfam; PF01600; Corona\_S1; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 76 76  
 SQ SEQUENCE 76 AA; 7861 MW; 9DA97501A9CB4FD1 CRC64;

Query Match 19.8%; Score 42; DB 12; Length 76;  
 Best Local Similarity 31.8%; Pred. No. 6.6e+02;  
 Matches 7; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 21 RVIENTPEALSVAVEEGLAWRK 42  
 |:: :|::| |::| |  
 Db 36 RIVNASSIAMTVPVGGQMOWSK 57

#### RESULT 22

Q9Z8X5

ID Q9Z8X5 PRELIMINARY; PRT; 79 AA.  
 AC Q9Z8X5;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein CPn0209.  
 GN CPN0209.  
 OS Chlamydia pneumoniae (Chlamydophila pneumoniae).  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CWL029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RL Nat. Genet. 21:385-389(1999).  
 DR EMBL; AE001607; AAD18362.1; -.  
 DR PIR; B72106; B72106.  
 DR InterPro; IPR006974; DUF648.  
 DR Pfam; PF04890; DUF648; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 79 AA; 9196 MW; 2813A36311D4A49A CRC64;

Query Match 19.8%; Score 42; DB 16; Length 79;  
 Best Local Similarity 33.3%; Pred. No. 6.9e+02;  
 Matches 8; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

Qy 15 FSGQKSRVIENPTEALSVAVEEGL 38  
 | |::||| |::| |::  
 Db 28 FQGKRTRVIAITPAGLAIAYEQNI 51

RESULT 23

Q9JSH8

ID Q9JSH8 PRELIMINARY; PRT; 79 AA.  
 AC Q9JSH8;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein CPJ0209.  
 GN CPJ0209.  
 OS Chlamydia pneumoniae (Chlamydophila pneumoniae).  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CWL029 from USA."  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 DR EMBL; AP002545; BAA98419.1; -.  
 DR PIR; A86517; A86517.  
 DR InterPro; IPR006974; DUF648.  
 DR Pfam; PF04890; DUF648; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 79 AA; 9212 MW; C70CA36311C3AFF7 CRC64;

Query Match 19.8%; Score 42; DB 16; Length 79;  
 Best Local Similarity 33.3%; Pred. No. 6.9e+02;  
 Matches 8; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

Qy 15 FSGQKSRVIENPTEALSVAVEEGL 38  
 | |::||| |::| |::  
 Db 28 FQGKRTRVIAITPAGLAIAYEQNI 51

RESULT 24

Q9K247

ID Q9K247 PRELIMINARY; PRT; 81 AA.  
 AC Q9K247;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein CP0557.  
 GN CP0557 OR CPB0213.  
 OS Chlamydia pneumoniae (Chlamydophila pneumoniae).  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;

RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TW-183;  
 RA Geng M.M., Schuhmacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,  
 RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;  
 RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with  
 RT other Chlamydia strains based on whole genome sequence analysis.";  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AE002214; AAF38377.1; -.  
 DR EMBL; AE017157; AAP98146.1; -.  
 DR PIR; D81565; D81565.  
 DR TIGR; CP0557; -.  
 DR InterPro; IPR006974; DUF648.  
 DR Pfam; PF04890; DUF648; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 81 AA; 9455 MW; C6A6483DA44594C2 CRC64;

Query Match 19.8%; Score 42; DB 16; Length 81;  
 Best Local Similarity 33.3%; Pred. No. 7.1e+02;  
 Matches 8; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

Qy 15 FSGQKSRVIENPTEALSVAVEEGL 38  
 | |::||| |::| |:::  
 Db 30 FQGRTRVIAITPAGLAIAYEQNI 53

# RESULT 25

Q88KZ7

ID Q88KZ7 PRELIMINARY; PRT; 77 AA.  
 AC Q88KZ7;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Conserved hypothetical protein.  
 GN PP2141.  
 OS Pseudomonas putida (strain KT2440).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=160488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22423060; PubMed=12534463;  
 RA Nelson K.E., Weinell C., Paulsen I.T., Dodson R.J., Hilbert H.,  
 RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,  
 RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,  
 RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,  
 RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzez A.,  
 RA Utterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,



RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,  
 RA Kiewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuemmler B.,  
 RA Fraser C.M.;  
 RT "Complete genome sequence and comparative analysis of the  
 RT metabolically versatile *Pseudomonas putida* KT2440.";  
 RL Environ. Microbiol. 4:799-808(2002).  
 DR EMBL; AE016782; AAN67754.1; -.  
 DR TIGR; PP2141; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 77 AA; 8556 MW; 861F24C53136CD4F CRC64;

Query Match 19.6%; Score 41.5; DB 16; Length 77;  
 Best Local Similarity 34.2%; Pred. No. 7.9e+02;  
 Matches 13; Conservative 6; Mismatches 14; Indels 5; Gaps 1;

QY 3 RSISENSLVAMDFSGQ-----KSRVIENPTEALSVAVE 35  
 | : | ||| : ||| : : : | ||| | :  
 Db 20 RAEDEGSLVTLDFSED AKVFLQGQHVEVAKAMLSVG VQ 57

Search completed: July 8, 2004, 08:22:50  
 Job time : 43.2441 secs

OM protein - protein search, using sw model

Run on: July 8, 2004, 08:03:43 ; Search time 13.5433 Seconds  
(without alignments)  
165.323 Million cell updates/sec

Title: US-09-936-697-5  
Perfect score: 212  
Sequence: 1 PMRSISENSLVAMDFSGQKS.....ENPTEALSVAVEEGLAWRKK 43

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 11046

Minimum DB seq length: 0  
Maximum DB seq length: 85

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	46.5	21.9	62	1	YZ05_METJA	Q60262 methanococc
2	46	21.7	73	1	RPON_METJA	Q57649 methanococc
3	42	19.8	82	1	Y567_METJA	Q57987 methanococc
4	41	19.3	79	1	DC13_HUMAN	Q9nnp2 homo sapien
5	40	18.9	82	1	RADC_STAAU	P31337 staphylococ
6	39.5	18.6	35	1	PBP_HYACE	P34175 hyalophora
7	39.5	18.6	60	1	Y574_LACLA	Q9chz4 lactococcus
8	39	18.4	45	1	RS22_ECOLI	P28690 escherichia
9	39	18.4	71	1	Y16K_BPT4	P39243 bacterioph
10	38.5	18.2	67	1	CSPF_STRCO	P48859 streptomyce
11	38.5	18.2	74	1	NIFH_NOSSN	P52336 nostoc sp.
12	38.5	18.2	78	1	CINA_STRGV	P29827 streptover
13	38	17.9	59	1	FER_METBA	P00202 methanosarc
14	38	17.9	60	1	YA87_STRMU	Q8du62 streptococc
15	38	17.9	63	1	FER2_DESVM	P10624 desulfovibr
16	38	17.9	72	1	RPON_THEAC	Q9hl09 thermoplasm
17	38	17.9	72	1	RPON_THEVO	Q979k0 thermoplasm

18	38	17.9	83	1	TMOB_PSEME	Q00457	pseudomonas
19	37	17.5	62	1	4OT_COMTE	Q9rh8	comamonas t
20	36	17.0	59	1	YH13_ARCFU	O28560	archaeoglob
21	36	17.0	62	1	4OT3_PSEPU	Q9z431	pseudomonas
22	36	17.0	62	1	4OT_PSEFL	Q8krr5	pseudomonas
23	36	17.0	66	1	CSP7_STRCL	Q01761	streptomyce
24	36	17.0	76	1	IPKG_HUMAN	Q9y2b9	homo sapien
25	36	17.0	76	1	IPKG_MOUSE	O70139	mus musculu
26	36	17.0	80	1	GCH1_MUCHA	P51598	mucuna hass
27	36	17.0	83	1	TRBG_ECOLI	P41072	escherichia
28	35.5	16.7	69	1	GBGU_BOVIN	P50154	bos taurus
29	35.5	16.7	75	1	ATP9_PARTE	P16001	paramecium
30	35	16.5	58	1	SINI_BACLI	P22755	bacillus li
31	35	16.5	67	1	HFOB_METFO	P48784	methanobact
32	35	16.5	70	1	ICIC_HIRME	P01051	hirudo medi
33	35	16.5	79	1	NSGX_HUMAN	Q9uh64	homo sapien
34	34.5	16.3	80	1	VPU_HVIMA	P05924	human immun
35	34	16.0	31	1	COG5_BOVIN	P83437	bos taurus
36	34	16.0	51	1	VG62_BPMD2	O64254	mycobacteri
37	34	16.0	59	1	SSH2_BACCR	Q81fb5	bacillus ce
38	34	16.0	60	1	Y2B4_BACHD	Q9k6b5	bacillus ha
39	34	16.0	60	1	YB39_STRPY	Q99zp7	streptococc
40	34	16.0	67	1	HFO2_METFO	P48783	methanobact
41	34	16.0	67	1	HMT1_METTH	P50483	methanobact
42	34	16.0	68	1	V07K_PMV	P20954	papaya mosa
43	34	16.0	78	1	Y8K2_SSV1	P20205	sulfolobus
44	34	16.0	79	1	YH65_SYNY3	P72913	synechocyst
45	33.5	15.8	60	1	CSRA_BUCBP	P59498	buchnera ap
46	33.5	15.8	82	1	YHYD_ANACY	P16420	anabaena cy
47	33	15.6	45	1	CYC6_PROHO	P81244	prochloroth
48	33	15.6	55	1	FER_BUTME	P14073	butyribacte
49	33	15.6	59	1	SSH2_BACAA	Q81sd1	bacillus an
50	33	15.6	59	1	YB11_STRA3	Q8e5c1	streptococc
51	33	15.6	75	1	YODD_ECOLI	P76328	escherichia
52	33	15.6	84	1	RL23_HALMA	P12732	haloarcula
53	32.5	15.3	45	1	VPU_HV1Z3	P08805	human immun
54	32.5	15.3	81	1	VPU_HV1JR	P20882	human immun
55	32.5	15.3	81	1	VPU_HV1S1	P19554	human immun
56	32	15.1	41	1	BAXC_HUMAN	Q07815	homo sapien
57	32	15.1	52	1	SRY_AKOAZ	P36388	akodon azar
58	32	15.1	60	1	NXS1_DENJA	P01417	dendroaspis
59	32	15.1	60	1	NXS1_DENV1	P01418	dendroaspis
60	32	15.1	60	1	YR09_LISIN	Q927s9	listeria in
61	32	15.1	62	1	4OT4_PSEPU	Q93jw0	pseudomonas
62	32	15.1	62	1	V452_BPT4	P07878	bacterioph
63	32	15.1	64	1	YEEW_ECOLI	P76366	escherichia
64	32	15.1	66	1	CADO_BUNCA	P81783	bungarus ca
65	32	15.1	66	1	RPON_SULSO	Q980z8	sulfolobus
66	32	15.1	67	1	YDFZ_ECOLI	P76153	escherichia
67	32	15.1	71	1	ACA1_ACALU	P81592	acalolepta
68	32	15.1	71	1	YQGC_ECOLI	P46878	escherichia
69	32	15.1	72	1	HAF1_ARCFU	O29910	archaeoglob
70	32	15.1	72	1	IF1_XANAC	Q8pl04	xanthomonas
71	32	15.1	72	1	IF1_XANCP	Q8p997	xanthomonas
72	32	15.1	73	1	VF14_VARV	P33872	variola vir
73	32	15.1	75	1	RPOK_ARCFU	O29134	archaeoglob
74	32	15.1	75	1	RPOZ_PORPU	P51376	porphyra pu

75	32	15.1	83	1	YCX1_PINTH	Q00865	pinus thunb
76	31.5	14.9	53	1	LECA_LATAP	P07441	lathyrus ap
77	31.5	14.9	57	1	V3A_IBVU4	P30239	avian infec
78	31.5	14.9	59	1	SSH1_BACAA	Q81v87	bacillus an
79	31.5	14.9	72	1	CCA2_ECOLI	Q46995	escherichia
80	31.5	14.9	74	1	YC65_LISIN	Q926b0	listeria in
81	31.5	14.9	79	1	Y124_AQUAE	O66524	aquifex aeo
82	31.5	14.9	81	1	YH25_XYLFA	Q9pcq3	xylella fas
83	31	14.6	25	1	ATPD_MICLU	P80285	micrococcus
84	31	14.6	43	1	CC3_CARCN	P32956	carica cand
85	31	14.6	43	1	CC4_CARCN	P32957	carica cand
86	31	14.6	46	1	V11_BPT3	P20833	bacterioph
87	31	14.6	55	1	RPON_METTH	O26147	methanobact
88	31	14.6	61	1	YWHB_BACSU	P70994	bacillus su
89	31	14.6	62	1	UCRX_BOVIN	P00130	bos taurus
90	31	14.6	63	1	JHEB_TRINI	P30810	trichoplusi
91	31	14.6	63	1	YAIA_ECOLI	P08366	escherichia
92	31	14.6	63	1	YDIE_ECOLI	P40721	escherichia
93	31	14.6	66	1	COPP_HELPJ	Q9zm70	helicobacte
94	31	14.6	66	1	DMS5_PACDA	O93455	pachymedusa
95	31	14.6	72	1	Y541_BORBU	O51491	borrelia bu
96	31	14.6	76	1	RPON_ARCFU	O29135	archaeoglob
97	31	14.6	77	1	RL28_HAEIN	P44364	haemophilus
98	31	14.6	78	1	YD80_METJA	Q58775	methanococc
99	31	14.6	80	1	AKA7_MOUSE	O55074	mus musculu
100	31	14.6	85	1	THIO_METJA	Q57755	methanococc

# ALIGNMENTS

## RESULT 1

YZ05\_METJA

ID YZ05\_METJA STANDARD; PRT; 62 AA.  
AC Q60262;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein MJECLO5.  
GN MJECLO5.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus

RT jannaschii.";  
 RL Science 273:1058-1073(1996).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L77118; AAC37071.1; -.  
 DR PIR; E64510; E64510.  
 DR TIGR; MJECL05; -.  
 KW Hypothetical protein; Complete proteome.  
 FT DOMAIN 3 15 ILE-RICH.  
 SQ SEQUENCE 62 AA; 7327 MW; 1624EC72E75EBAD7 CRC64;

Query Match 21.9%; Score 46.5; DB 1; Length 62;  
 Best Local Similarity 28.6%; Pred. No. 19;  
 Matches 12; Conservative 8; Mismatches 21; Indels 1; Gaps 1;

Qy 3 RSISENSLVAMDFS-GQKSRVIENPTEALSVAVEEGLAWRKK 43  
 : ::| | ::| | : | | :|||: | ||  
 Db 18 KKVAERFLKDLESSQGM DWKEIRERAERAKKQLEEGIEWAKK 59

## RESULT 2

### RPON\_METJA

ID RPON\_METJA STANDARD; PRT; 73 AA.  
 AC Q57649;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE DNA-directed RNA polymerase subunit N (EC 2.7.7.6).  
 GN RPON OR MJ0196.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RT jannaschii.";  
 RL Science 273:1058-1073(1996).  
 CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription  
 CC of DNA into RNA using the four ribonucleoside triphosphates as

CC substrates.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA} (N).  
 CC -!- SIMILARITY: Belongs to the archaeal rpoN / eukaryotic RPB10 RNA  
 CC polymerase subunit family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U67475; AAB98176.1; -.  
 DR HSSP; O26147; 1EF4.  
 DR TIGR; MJ0196; -.  
 DR HAMAP; MF\_00250; -; 1.  
 DR InterPro; IPR000268; RNA\_pol\_N.  
 DR Pfam; PF01194; RNA\_pol\_N; 1.  
 DR ProDom; PD006539; RNA\_pol\_N; 1.  
 DR PROSITE; PS01112; RNA\_POL\_N\_8KD; 1.  
 KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc;  
 KW Metal-binding; Complete proteome.  
 FT METAL 7 7 ZINC (BY SIMILARITY).  
 FT METAL 10 10 ZINC (BY SIMILARITY).  
 FT METAL 44 44 ZINC (BY SIMILARITY).  
 FT METAL 45 45 ZINC (BY SIMILARITY).  
 SQ SEQUENCE 73 AA; 8695 MW; E716EA406D65B831 CRC64;

Query Match 21.7%; Score 46; DB 1; Length 73;  
 Best Local Similarity 34.4%; Pred. No. 26;  
 Matches 11; Conservative 7; Mismatches 12; Indels 2; Gaps 1;

Qy 1 PMRSISENSIVAMDFSGQKSRVI--ENPTEAL 30  
 |:| | :::| | | |:: ||| : |  
 Db 4 PIRCFSCGNVIAEVFEEYKERILKGENPKDVL 35

# RESULT 3

Y567\_METJA

ID Y567\_METJA STANDARD; PRT; 82 AA.  
 AC Q57987;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MJ0567.  
 GN MJ0567.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,

RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, *Methanococcus*  
 RT *jannaschii*.";  
 RL Science 273:1058-1073(1996).

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; U67505; AAB98558.1; -.  
 DR PIR; G64370; G64370.  
 DR TIGR; MJ0567; -.  
 DR InterPro; IPR007167; FeoA.  
 DR InterPro; IPR008988; Transcr\_rep\_C.  
 DR Pfam; PF04023; FeoA; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 82 AA; 8766 MW; 3F3810EEFC9F81CE CRC64;

Query Match 19.8%; Score 42; DB 1; Length 82;  
 Best Local Similarity 32.5%; Pred. No. 1e+02;  
 Matches 13; Conservative 8; Mismatches 11; Indels 8; Gaps 2;

Qy 10 LVAMDFS-GQKSRVIEN-----PTEALSVAVEEGLAWR 41  
 ||:| :| | :|| | | : :||: ||| :  
 Db 28 LVSMGINIGSKLKVIRNQNGPVIISTKGSNIAIGRGLAMK 67

#### RESULT 4

##### DC13\_HUMAN

ID DC13\_HUMAN STANDARD; PRT; 79 AA.  
 AC Q9NRP2;  
 DT 15-MAR-2004 (Rel. 43, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE UPF0287 protein DC13.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Dendritic cell;  
 RA Gu Y., Peng Y., Li N., Gu W., Han Z., Fu G., Chen Z.;  
 RT "Novel genes expressed in human dendritic cells.";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC TISSUE=Breast;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -!- SIMILARITY: Belongs to the UPF0287 family.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AF201935; AAF86871.1; -.  
DR EMBL; BC032631; AAH32631.1; -.  
SQ SEQUENCE 79 AA; 9460 MW; 783381BD6DAFB7AA CRC64;

Query Match 19.3%; Score 41; DB 1; Length 79;  
Best Local Similarity 40.0%; Pred. No. 1.3e+02;  
Matches 10; Conservative 5; Mismatches 6; Indels 4; Gaps 1;

Qy 19 KSRVIENPTEALSVAVEEGLAWRKK 43  
|: :||| |: | |:| |||  
Db 49 KNEYVENRTKSR----EHGIAMRKK 69

RESULT 5  
RADC\_STAAU  
ID RADC\_STAAU STANDARD; PRT; 82 AA.  
AC P31337;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE DNA repair protein radC homolog (25 kDa protein) (Fragment).  
GN RADC.  
OS Staphylococcus aureus.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1280;



```

RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=RN450;
RA  Murphy E.;
RL  Submitted (JAN-1986) to the EMBL/GenBank/DDBJ databases.
RN  [2]
RP  PARTIAL SEQUENCE FROM N.A.
RC  STRAIN=RN450;
RX  MEDLINE=84117462; PubMed=6320000;
RA  Murphy E., Loefdahl S.;
RT  "Transposition of Tn554 does not generate a target duplication.";
RL  Nature 307:292-294(1984).
CC  -!- FUNCTION: Involved in DNA repair (By similarity).
CC  -!- SIMILARITY: Belongs to the radC family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; K02985; AAA26680.1; -.
DR  HAMAP; MF_00018; -; 1.
DR  InterPro; IPR001405; RadC.
DR  Pfam; PF04002; RadC; 1.
DR  ProDom; PD007415; RadC; 1.
DR  PROSITE; PS01302; RADC; 1.
KW  DNA repair.
FT  NON_TER      1      1
SQ  SEQUENCE    82 AA;  8920 MW;  65E8BF06E3DEC3A4 CRC64;

```

```

Query Match          18.9%; Score 40; DB 1; Length 82;
Best Local Similarity 40.9%; Pred. No. 1.9e+02;
Matches      9; Conservative      3; Mismatches    10; Indels      0; Gaps      0;

```

```

Qy      15 FSGQKSRVIENPTEALSVAVEE 36
      | | : | : | | | : | | |
Db      1 FKGTLNSSIVHPREIFSIIVRE 22

```

# RESULT 6

```

PBP_HYACE
ID  PBP_HYACE          STANDARD;          PRT;          35 AA.
AC  P34175;
DT  01-FEB-1994 (Rel. 28, Created)
DT  01-FEB-1994 (Rel. 28, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Pheromone-binding protein (PBP) (Fragment).
OS  Hyalophora cecropia (Cecropia moth).
OC  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC  Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
OC  Saturniidae; Saturniinae; Attacini; Hyalophora.
OX  NCBI_TaxID=7123;
RN  [1]
RP  SEQUENCE.

```

RX MEDLINE=91186129; PubMed=2010751;  
 RA Vogt R.G., Prestwich G.D., Lerner M.R.;  
 RT "Odorant-binding-protein subfamilies associate with distinct classes  
 RT of olfactory receptor neurons in insects.";  
 RL J. Neurobiol. 22:74-84(1991).  
 CC -!- FUNCTION: THIS MAJOR SOLUBLE PROTEIN IN OLFACTORY SENSILLA OF MALE  
 CC MOTHS MIGHT SERVE TO SOLUBILIZE THE EXTREMELY HYDROPHOBIC  
 CC PHEROMONE MOLECULES AND TO TRANSPORT PHEROMONE THROUGH THE AQUEOUS  
 CC LYMPH TO RECEPTORS LOCATED ON OLFACTORY CILIA.  
 CC -!- TISSUE SPECIFICITY: Antenna.  
 CC -!- SIMILARITY: Belongs to the PBP/GOBP family.  
 DR HSSP; P34174; 1DQE.  
 DR InterPro; IPR006170; PBP\_GOBP.  
 DR Pfam; PF01395; PBP\_GOBP; 1.  
 KW Pheromone-binding; Pheromone response; Transport.  
 FT NON\_TER 35 35  
 SQ SEQUENCE 35 AA; 4061 MW; 9B1B9D20D472E769 CRC64;

Query Match 18.6%; Score 39.5; DB 1; Length 35;  
 Best Local Similarity 37.9%; Pred. No. 85;  
 Matches 11; Conservative 5; Mismatches 10; Indels 3; Gaps 1;

Qy 2 MRSISENSLVAMDFSGQKSRVIENPTEAL 30  
 |:|:| | | | : : | :  
 Db 5 MKSLSENFCAMD---QCKQELNLPDEVI 30

#### RESULT 7

Y574\_LACLA

ID Y574\_LACLA STANDARD; PRT; 60 AA.  
 AC Q9CHZ4;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Probable tautomerase LL0574 (EC 5.3.2.-).  
 GN LL0574.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
 OX NCBI\_TaxID=1360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IL1403;  
 RX MEDLINE=21235186; PubMed=11337471;  
 RA Bolotin A., Wincker P., Mager S., Jaillon O., Malarme K.,  
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;  
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
 RT lactis ssp. lactis IL1403.";  
 RL Genome Res. 11:731-753(2001).  
 CC -!- SIMILARITY: Belongs to the tautomerase family.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL; AE006291; AAK04672.1; -.
DR PIR; F86696; F86696.
DR HAMAP; MF_00718; -; 1.
DR InterPro; IPR004370; Taut.
DR Pfam; PF01361; Tautomerase; 1.
DR ProDom; PD404143; Taut; 1.
KW Isomerase; Complete proteome.
FT INIT_MET      0      0      BY SIMILARITY.
FT ACT_SITE      1      1      CATALYTIC BASE (BY SIMILARITY).
SQ SEQUENCE      60 AA;  6667 MW;  19E80C7BA3EAFFFF CRC64;

```

```

Query Match          18.6%; Score 39.5; DB 1; Length 60;
Best Local Similarity 21.4%; Pred. No. 1.5e+02;
Matches      9; Conservative 14; Mismatches 16; Indels 3; Gaps 1;

```

```

QY      3 RSISENSLVAMDFSGQKSRVIENPTEALSVA---VEEGLAWR 41
      |:: : ::| : : | : || : | : ||: ::
Db      11 RTVEQKAIIAKEITESISKHAGAPSAIHVIFNDLPEGMLYQ 52

```

# RESULT 8

## RS22\_ECOLI

```

ID RS22_ECOLI      STANDARD;      PRT;      45 AA.
AC P28690;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S22 (Stationary-phase-induced ribosome-
DE associated protein) (SRA) (Protein D).
GN RPSV OR SRA OR B1480 OR C1913 OR Z2230 OR ECS2084.
OS Escherichia coli,
OS Escherichia coli O6, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992, 83334;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-37.
RC STRAIN=K12 / W3110;
RX MEDLINE=21189300; PubMed=11292794;
RA Izutsu K., Wada C., Komine Y., Sako T., Ueguchi C., Nakura S.,
RA Wada A.;
RT "Escherichia coli ribosome-associated protein SRA, whose copy number
RT increases during stationary phase.";
RL J. Bacteriol. 183:2765-2773(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90337272; PubMed=2199308;
RA Mahajan S.K., Chu C.C., Willis D.K., Templin A., Clark A.J.;
RT "Physical analysis of spontaneous and mutagen-induced mutants of
RT Escherichia coli K-12 expressing DNA exonuclease VIII activity.";
RL Genetics 125:261-273(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;

```

RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12."  
 RL Science 277:1453-1474(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic Escherichia coli."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."  
 RL Nature 409:529-533(2001).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12."  
 RL DNA Res. 8:11-22(2001).  
 RN [7]  
 RP MASS SPECTROMETRY.  
 RC STRAIN=K12 / ATCC 25404;  
 RX MEDLINE=99196679; PubMed=10094780;  
 RA Arnold R.J., Reilly J.P.;  
 RT "Observation of Escherichia coli ribosomal proteins and their  
 RT posttranslational modifications by mass spectrometry."  
 RL Anal. Biochem. 269:105-112(1999).  
 CC -!- MASS SPECTROMETRY: MW=5095.9; METHOD=MALDI.  
 CC -!- SIMILARITY: Belongs to the S22P family of ribosomal proteins.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; D13179; BAA02474.1; -.  
DR EMBL; X55956; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; AE000245; AAC74553.1; -.  
DR EMBL; AE016760; AAN80372.1; -.  
DR EMBL; AE005357; AAG56289.1; -.  
DR EMBL; AP002557; BAB35507.1; -.  
DR PIR; C64901; C64901.  
DR PIR; D90889; D90889.  
DR PIR; E85728; E85728.  
DR EcoGene; EG11508; rpsV.  
KW Ribosomal protein; Complete proteome.  
SQ SEQUENCE 45 AA; 5096 MW; 81DB6E2D2E222C2F CRC64;

Query Match 18.4%; Score 39; DB 1; Length 45;  
Best Local Similarity 63.6%; Pred. No. 1.3e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 17 GQKSRVIENPT 27  
| || |: |||  
Db 27 GDKSSVVNNPT 37

#### RESULT 9

Y16K\_BPT4

ID Y16K\_BPT4 STANDARD; PRT; 71 AA.  
AC P39243;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hypothetical 8.1 kDa protein in ndd-denB intergenic region.  
GN Y16K OR NDD.1.  
OS Bacteriophage T4.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
OC T4-like viruses.  
OX NCBI\_TaxID=10665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22514363; PubMed=12626685;  
RA Miller E.S., Kutter E., Mosig G., Arisaka F., Kunisawa T., Ruger W.;  
RT "Bacteriophage T4 genome.";  
RL Microbiol. Mol. Biol. Rev. 67:86-156(2003).  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AF158101; AAD42616.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 71 AA; 8143 MW; 5D56546D2FADAF0C CRC64;

Query Match 18.4%; Score 39; DB 1; Length 71;  
Best Local Similarity 32.4%; Pred. No. 2.2e+02;  
Matches 11; Conservative 5; Mismatches 16; Indels 2; Gaps 1;

Qy 1 PMRSISENSLVAMDFSGQKSR--VIENPTEALSV 32  
|::| || | : : ||| || :|  
Db 26 PLKSTSEKMTVNATLANNSNERFCIENDTETYTV 59

RESULT 10

CSPF\_STRCO

ID CSPF\_STRCO STANDARD; PRT; 67 AA.  
AC P48859;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cold shock protein scoF.  
GN SCOF OR SCO0527 OR SCF11.07C.  
OS Streptomyces coelicolor.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Av-Gay Y., Ravin S., Aharonowitz Y., Cohen G.;  
RL Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RX MEDLINE=21996410; PubMed=12000953;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
RA Rabbिनowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
RT coelicolor A3(2).";  
RL Nature 417:141-147(2002).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
CC -!- INDUCTION: In response to low temperature.  
CC -!- SIMILARITY: Belongs to the cold-shock domain (CSD) family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X92686; CAA63367.1; -.  
DR EMBL; AL939105; CAB59584.1; -.

DR PIR; T42055; T42055.  
 DR HSSP; P32081; 1CSP.  
 DR InterPro; IPR002059; Cold\_shock.  
 DR InterPro; IPR008994; Nucleic\_acid\_OB.  
 DR Pfam; PF00313; CSD; 1.  
 DR PRINTS; PR00050; COLDSHOCK.  
 DR ProDom; PD000621; Cold\_shock; 1.  
 DR SMART; SM00357; CSP; 1.  
 DR PROSITE; PS00352; COLD\_SHOCK; 1.  
 KW Transcription regulation; DNA-binding; Activator; Complete proteome.  
 FT DOMAIN 4 64 CSD.  
 SQ SEQUENCE 67 AA; 7179 MW; E4FDAD9BB1D92B34 CRC64;

Query Match 18.2%; Score 38.5; DB 1; Length 67;  
 Best Local Similarity 39.3%; Pred. No. 2.4e+02;  
 Matches 11; Conservative 2; Mismatches 14; Indels 1; Gaps 1;

QY 3 RSISENSLVAMDFS-GQKSRVIENPTEA 29  
 | : | | : || | |  
 Db 40 RELQEGQAVTFDITQGQKGPQAENITPA 67

# RESULT 11

## NIFH\_NOSSN

ID NIFH\_NOSSN STANDARD; PRT; 74 AA.  
 AC P52336;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Nitrogenase iron protein (EC 1.18.6.1) (Nitrogenase component II)  
 DE (Nitrogenase Fe protein) (Nitrogenase reductase) (Fragment).  
 GN NIFH.  
 OS Nostoc sp. (strain MUN 8820).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCBI\_TaxID=55397;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97086627; PubMed=8932316;  
 RA Hill D.R., Belbin T.J., Thorsteinsson M.V., Bassam D., Brass S.,  
 RA Ernst A., Boger P., Paerl H., Mulligan M.E., Potts M.;  
 RT "GlbN (cyanoglobin) is a peripheral membrane protein that is  
 RT restricted to certain Nostoc spp."  
 RL J. Bacteriol. 178:6587-6598(1996).  
 CC -!- FUNCTION: The key enzymatic reactions in nitrogen fixation are  
 CC catalyzed by the nitrogenase complex, which has 2 components: the  
 CC iron protein and the molybdenum-iron protein.  
 CC -!- CATALYTIC ACTIVITY: 8 reduced ferredoxin + 8 H(+) + N(2) + 16 ATP  
 CC = 8 oxidized ferredoxin + 2 NH(3) + 16 ADP + 16 phosphate.  
 CC -!- COFACTOR: Binds one 4Fe-4S cluster per dimer.  
 CC -!- SUBUNIT: Homodimer (By similarity).  
 CC -!- SIMILARITY: Belongs to the nifH / bchL / chlL family.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; L47979; AAB41123.1; -.  
DR HSSP; P00459; 1FP6.  
DR HAMAP; MF\_00533; -; 1.  
DR InterPro; IPR000392; NitrogenaseII.  
DR Pfam; PF00142; fer4\_NifH; 1.  
DR PRINTS; PR00091; NITROGNASEII.  
DR PROSITE; PS00692; NIFH\_FRXC\_2; PARTIAL.  
DR PROSITE; PS00746; NIFH\_FRXC\_1; PARTIAL.  
KW Oxidoreductase; Nitrogen fixation; Iron-sulfur; 4Fe-4S; ATP-binding.  
FT NP\_BIND 13 20 ATP (POTENTIAL).  
FT NON\_TER 74 74  
SQ SEQUENCE 74 AA; 7919 MW; 14B88F560242DCDE CRC64;

Query Match 18.2%; Score 38.5; DB 1; Length 74;  
Best Local Similarity 25.5%; Pred. No. 2.6e+02;  
Matches 12; Conservative 7; Mismatches 13; Indels 15; Gaps 1;

Qy 6 SENSLVAMDFSGQKSRVI-----ENPTEALSVAVEEG 37  
|:|:| || ||: :: : | | :| || |  
Db 23 SQNTLAAMAEMGQRILIVGCDPKADSTRMLHLSKAQTTVLHLAAERG 69

#### RESULT 12

##### CINA\_STRGV

ID CINA\_STRGV STANDARD; PRT; 78 AA.  
AC P29827;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Lantibiotic cinnamycin precursor (Lanthiopeptin) (Lantibiotic Ro  
DE 09-0198).  
GN CINA OR ROCA.  
OS Streptovercillium griseovercillatum.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=68215;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAR 164C-MY6;  
RX MEDLINE=91301152; PubMed=2070795;  
RA Kaletta C., Entian K.-D., Jung G.;  
RT "Prepeptide sequence of cinnamycin (Ro 09-0198): the first structural  
RT gene of a duramycin-type lantibiotic.";  
RL Eur. J. Biochem. 199;411-415(1991).  
RN [2]  
RP SEQUENCE OF 60-78.  
RX MEDLINE=91107436; PubMed=2125590;  
RA Fredenhagen A., Fendrich G., Marki F., Marki W., Gruner J.,  
RA Raschdorf F., Peter H.H.;  
RT "Duramycins B and C, two new lanthionine containing antibiotics as  
RT inhibitors of phospholipase A2. Structural revision of duramycin and  
RT cinnamycin.";  
RL J. Antibiot. 43:1403-1412(1990).  
RN [3]



RP SEQUENCE OF 60-78.  
RX MEDLINE=89291558; PubMed=2544544;  
RA Naruse N., Tenmyo O., Tomita K., Konishi M., Miyaki T., Kawaguchi H.,  
RA Fukase K., Wakamiya T., Shiba T.;  
RT "Lanthiopeptin, a new peptide antibiotic. Production, isolation and  
RT properties of lanthiopeptin.";  
RL J. Antibiot. 42:837-845(1989).  
CC -!- FUNCTION: Can act as inhibitor of the enzyme phospholipase A2, and  
CC of the angiotensin-converting enzyme. Shows inhibitory activities  
CC against herpes simplex virus and immunopotentiating activities.  
CC Its antimicrobial activities are not very pronounced.  
CC -!- PTM: Maturation of lantibiotics involves the enzymic conversion of  
CC Thr, and Ser into dehydrated AA and the formation of thioether  
CC bonds with cysteine or the formation of dialkylamine bonds with  
CC lysine. This is followed by membrane translocation and cleavage of  
CC the modified precursor.  
CC -!- SIMILARITY: Belongs to the type B lantibiotic family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X58545; CAA41436.1; -.  
DR PIR; A45767; EWSMCN.  
DR PIR; S17181; EWSMYG.  
KW Antibiotic; Bacteriocin; Lantibiotic; Thioether bond.  
FT PROPEP 1 59 POTENTIAL.  
FT CHAIN 60 78 LANTIBIOTIC CINNAMYCIN.  
FT CROSSLNK 60 77 Beta-methyllanthionine (Cys-Thr).  
FT CROSSLNK 63 73 Lanthionine (Ser-Cys).  
FT CROSSLNK 64 70 Beta-methyllanthionine (Cys-Thr).  
FT CROSSLNK 65 78 Lysinoalanine (Ser-Lys).  
SQ SEQUENCE 78 AA; 8205 MW; 0ACDAE6BA54E5E7A CRC64;

Query Match 18.2%; Score 38.5; DB 1; Length 78;  
Best Local Similarity 34.2%; Pred. No. 2.8e+02;  
Matches 13; Conservative 6; Mismatches 10; Indels 9; Gaps 2;

Qy 4 SISENSLVAMDFSGQKSRVIENP-----TEALSVAVE 35  
|| : |:| || :: :||| || ||  
Db 4 SILQQSVVDADF---RAALLENPAAFGASAAALPTPVE 38

# RESULT 13

## FER METBA

ID FER METBA STANDARD; PRT; 59 AA.  
AC P00202;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ferredoxin.  
OS Methanosarcina barkeri.  
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;

OC Methanosarcinaceae; Methanosarcina.  
 OX NCBI\_TaxID=2208;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=MS / DSM 800;  
 RX MEDLINE=83056954; PubMed=6754724;  
 RA Hausinger R.P., Moura I., Moura J.J.G., Xavier A.V., Santos M.H.,  
 RA Legall J., Howard J.B.;  
 RT "Amino acid sequence of a 3Fe:3S ferredoxin from the  
 RT 'archaeobacterium' Methanosarcina barkeri (DSM 800).";  
 RL J. Biol. Chem. 257:14192-14197(1982).  
 CC -!- FUNCTION: Ferredoxins are iron-sulfur proteins that transfer  
 CC electrons in a wide variety of metabolic reactions.  
 CC -!- COFACTOR: Binds 2 4Fe-4S clusters.  
 CC -!- SIMILARITY: Belongs to the bacterial-type ferredoxin family.  
 DR PIR; A00204; FEMZB.  
 DR HSSP; P00214; 2FD2.  
 DR InterPro; IPR001450; 4Fe4S\_ferredoxin.  
 DR InterPro; IPR000813; 7Fe\_ferredoxin.  
 DR Pfam; PF00037; fer4; 2.  
 DR PRINTS; PR00354; 7FE8SFRDOXIN.  
 DR PROSITE; PS00198; 4FE4S\_FERREDOXIN; 2.  
 KW Electron transport; Iron-sulfur; 4Fe-4S.  
 FT METAL 9 9 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).  
 FT METAL 12 12 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).  
 FT METAL 15 15 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).  
 FT METAL 19 19 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).  
 FT METAL 40 40 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).  
 FT METAL 43 43 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).  
 FT METAL 46 46 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).  
 FT METAL 50 50 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).  
 SQ SEQUENCE 59 AA; 6121 MW; 22D1EB8E443422CA CRC64;  
  
 Query Match 17.9%; Score 38; DB 1; Length 59;  
 Best Local Similarity 35.7%; Pred. No. 2.4e+02;  
 Matches 10; Conservative 8; Mismatches 10; Indels 0; Gaps 0;  
  
 QY 12 AMDFSGQKSRVIENPTEALSVAVEEGLA 39  
 | : || : | | | :|:: | :| :|  
 Db 6 ADECSGCGTCVDECPNDAITLDEEKGIA 33

# RESULT 14

YA87\_STRMU  
 ID YA87\_STRMU STANDARD; PRT; 60 AA.  
 AC Q8DU62;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Probable tautomerase SMU.1087 (EC 5.3.2.-).  
 GN SMU.1087.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=UA159 / ATCC 700610 / Serotype C;  
 RX MEDLINE=22295063; PubMed=12397186;  
 RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,  
 RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;  
 RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental  
 RT pathogen.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).  
 CC -!- SIMILARITY: Belongs to the tautomerase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE014946; AAN58785.1; -.  
 DR HAMAP; MF\_00718; -; 1.  
 DR InterPro; IPR004370; Taut.  
 DR Pfam; PF01361; Tautomerase; 1.  
 DR ProDom; PD404143; Taut; 1.  
 KW Isomerase; Complete proteome.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT ACT\_SITE 1 1 CATALYTIC BASE (BY SIMILARITY).  
 SQ SEQUENCE 60 AA; 6872 MW; 0ADFFDF5985622F4 CRC64;

Query Match 17.9%; Score 38; DB 1; Length 60;  
 Best Local Similarity 29.4%; Pred. No. 2.4e+02;  
 Matches 10; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

Qy 3 RSISENSLVAMDFSGQKSRVIENPTEALSVAVEE 36  
 || : :| : : ||| : | ||: | : :  
 Db 11 RSQEQKIQLAREVTEVVSRVAKAPKEAIHV FIND 44

# RESULT 15

## FER2\_DESVM

ID FER2\_DESVM STANDARD; PRT; 63 AA.  
 AC P10624;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ferredoxin II (Fd II).  
 OS Desulfovibrio vulgaris (strain Miyazaki).  
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;  
 OC Desulfovibrionaceae; Desulfovibrio.  
 OX NCBI\_TaxID=883;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kitamura M., Konishi T., Kawanishi K., Ohashi K., Kishida M.,  
 RA Kohno K., Akutsu H., Kumagai I., Nakaya T.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE.  
 RX MEDLINE=89274328; PubMed=2855025;

RA Okawara N., Ogata M., Yagi T., Wakabayashi S., Matsubara H.;  
 RT "Characterization and complete amino acid sequence of ferredoxin II  
 RT from *Desulfovibrio vulgaris* Miyazaki.";  
 RL Biochimie 70:1815-1820(1988).  
 CC -!- FUNCTION: Ferredoxins are iron-sulfur proteins that transfer  
 CC electrons in a wide variety of metabolic reactions.  
 CC -!- COFACTOR: Binds 1 4Fe-4S cluster.  
 CC -!- SUBUNIT: Homodimer.  
 CC -!- SIMILARITY: Belongs to the bacterial-type ferredoxin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB005550; BAA21477.1; -.  
 DR PIR; S07154; FEDV2V.  
 DR HSSP; P00210; 1FXR.  
 DR InterPro; IPR001080; 3Fe4S\_ferredoxin.  
 DR InterPro; IPR001450; 4Fe4S\_ferredoxin.  
 DR Pfam; PF00037; fer4; 1.  
 DR PRINTS; PR00352; 3FE4SFRDOXIN.  
 DR PROSITE; PS00198; 4FE4S\_FERREDOXIN; 1.  
 KW Electron transport; Iron-sulfur; Repeat; 4Fe-4S.  
 FT INIT\_MET 0 0  
 FT METAL 11 11 IRON-SULFUR (4FE-4S).  
 FT METAL 14 14 IRON-SULFUR (4FE-4S).  
 FT METAL 17 17 IRON-SULFUR (4FE-4S).  
 FT METAL 53 53 IRON-SULFUR (4FE-4S).  
 SQ SEQUENCE 63 AA; 7091 MW; 82232C1244A5C84B CRC64;

Query Match 17.9%; Score 38; DB 1; Length 63;  
 Best Local Similarity 27.0%; Pred. No. 2.6e+02;  
 Matches 10; Conservative 9; Mismatches 12; Indels 6; Gaps 1;

Qy 13 MDFSGQKSRVIENPT-----EALSVAVEEGLAWRKK 43  
 | :|: :||: | :|:| | :||::  
 Db 27 MSSAGEYAEVIDPNTTAECEVEDAISTCPVECI EWREE 63

#### RESULT 16

##### RPON\_THEAC

ID RPON\_THEAC STANDARD; PRT; 72 AA.  
 AC Q9HL09;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE DNA-directed RNA polymerase subunit N (EC 2.7.7.6).  
 GN RPON OR TA0431.  
 OS *Thermoplasma acidophilum*.  
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;  
 OC Thermoplasmataceae; Thermoplasma.  
 OX NCBI\_TaxID=2303;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM 1728;  
 RX MEDLINE=20479972; PubMed=11029001;  
 RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
 RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;  
 RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma  
 RT acidophilum.";  
 RL Nature 407:508-513(2000).  
 CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription  
 CC of DNA into RNA using the four ribonucleoside triphosphates as  
 CC substrates.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA}(N).  
 CC -!- SIMILARITY: Belongs to the archaeal rpoN / eukaryotic RPB10 RNA  
 CC polymerase subunit family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL445064; CAC11573.1; -.  
 DR HSSP; O26147; 1EF4.  
 DR HAMAP; MF\_00250; -; 1.  
 DR InterPro; IPR000268; RNA\_pol\_N.  
 DR Pfam; PF01194; RNA\_pol\_N; 1.  
 DR ProDom; PD006539; RNA\_pol\_N; 1.  
 DR PROSITE; PS01112; RNA\_POL\_N\_8KD; 1.  
 KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc;  
 KW Metal-binding; Complete proteome.  
 FT METAL 7 7 ZINC (BY SIMILARITY).  
 FT METAL 10 10 ZINC (BY SIMILARITY).  
 FT METAL 53 53 ZINC (BY SIMILARITY).  
 FT METAL 54 54 ZINC (BY SIMILARITY).  
 SQ SEQUENCE 72 AA; 8368 MW; 792AEDA20E5447E2 CRC64;

Query Match 17.9%; Score 38; DB 1; Length 72;  
 Best Local Similarity 30.6%; Pred. No. 3e+02;  
 Matches 11; Conservative 5; Mismatches 20; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEE 36  
 |:| | ::| |: || | | :||  
 Db 4 PVRCFSCGRVIASDYGRYIKRVNEIKAEGRDPSPEE 39

#### RESULT 17

RPON\_THEVO

ID RPON\_THEVO STANDARD; PRT; 72 AA.  
 AC Q979K0;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE DNA-directed RNA polymerase subunit N (EC 2.7.7.6).  
 GN RPON OR TV1161 OR TVG1188103.

OS Thermoplasma volcanium.  
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;  
 OC Thermoplasmataceae; Thermoplasma.  
 OX NCBI\_TaxID=50339;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GSS1 / DSM 4299 / JCM 9571;  
 RX MEDLINE=20570466; PubMed=11121031;  
 RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,  
 RA Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,  
 RA Nunoshiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;  
 RT "Archaeal adaptation to higher temperatures revealed by genomic  
 RT sequence of Thermoplasma volcanium."  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).  
 CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription  
 CC of DNA into RNA using the four ribonucleoside triphosphates as  
 CC substrates.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA} (N).  
 CC -!- SIMILARITY: Belongs to the archaeal rpoN / eukaryotic RPB10 RNA  
 CC polymerase subunit family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AP000995; BAB60303.1; -.  
 DR HAMAP; MF\_00250; -; 1.  
 DR InterPro; IPR000268; RNA\_pol\_N.  
 DR Pfam; PF01194; RNA\_pol\_N; 1.  
 DR ProDom; PD006539; RNA\_pol\_N; 1.  
 DR PROSITE; PS01112; RNA\_POL\_N\_8KD; 1.  
 KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc;  
 KW Metal-binding; Complete proteome.  
 FT METAL 7 7 ZINC (BY SIMILARITY).  
 FT METAL 10 10 ZINC (BY SIMILARITY).  
 FT METAL 53 53 ZINC (BY SIMILARITY).  
 FT METAL 54 54 ZINC (BY SIMILARITY).  
 SQ SEQUENCE 72 AA; 8483 MW; 06AEC0AA7AC75CA6 CRC64;

Query Match 17.9%; Score 38; DB 1; Length 72;  
 Best Local Similarity 27.8%; Pred. No. 3e+02;  
 Matches 10; Conservative 6; Mismatches 20; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEE 36  
 |:| | ::| |: | :| ||  
 Db 4 PVRCFSCGRVIASDYGRYLRRINEIRSEGREPTAEE 39

RESULT 18  
 TMOB\_PSEME  
 ID TMOB\_PSEME STANDARD; PRT; 83 AA.  
 AC Q00457;

DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Toluene-4-monooxygenase system protein B (EC 1.14.13.-).  
 GN TMOB.  
 OS Pseudomonas mendocina.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=300;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-27.  
 RC STRAIN=KR1;  
 RX MEDLINE=91358306; PubMed=1885512;  
 RA Yen K.-M., Karl M.R., Blatt L.M., Simon M.J., Winter R.B.,  
 RA Fausset P.R., Lu H.S., Harcourt A.A., Chen K.K.;  
 RT "Cloning and characterization of a Pseudomonas mendocina KR1 gene  
 RT cluster encoding toluene-4-monooxygenase.";  
 RL J. Bacteriol. 173:5315-5327(1991).  
 CC -!- FUNCTION: HYDROXYLATES TOLUENE TO FORM P-CRESOL.  
 CC -!- COFACTOR: FAD; requires Fe(2+) for activity.  
 CC -!- PATHWAY: Toluene degradation; first step.  
 CC -!- SUBUNIT: THE MULTICOMPONENT ENZYME TOLUENE-4-MONOOXYGENASE  
 CC IS FORMED BY THE TMOA, TMOB, TMOC, TMOD, TMOE AND TMOF  
 CC POLYPEPTIDES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M65106; AAA26000.1; -.  
 KW Aromatic hydrocarbons catabolism; Oxidoreductase; Flavoprotein;  
 KW Monooxygenase; FAD; Iron.  
 FT INIT\_MET 0 0  
 SQ SEQUENCE 83 AA; 9457 MW; 4729FEF73F266F44 CRC64;

Query Match 17.9%; Score 38; DB 1; Length 83;  
 Best Local Similarity 58.3%; Pred. No. 3.5e+02;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 25 NPTEALSVAVEE 36  
 |||| : | ||  
 Db 72 NPTEVIDVVFEE 83

#### RESULT 19

4OT\_COMTE

ID 4OT\_COMTE STANDARD; PRT; 62 AA.  
 AC Q9RHM8;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE 4-oxalocrotonate tautomerase (EC 5.3.2.-) (4-OT).  
 GN APhi.

OS Comamonas testosteroni (Pseudomonas testosteroni).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Comamonadaceae; Comamonas.  
 OX NCBI\_TaxID=285;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TA441;  
 RX MEDLINE=20340973; PubMed=10878134;  
 RA Arai H., Ohishi T., Chang M.Y., Kudo T.;  
 RT "Arrangement and regulation of the genes for meta-pathway enzymes  
 RT required for degradation of phenol in Comamonas testosteroni TA441.";  
 RL Microbiology 146:1707-1715(2000).  
 CC -!- FUNCTION: Catalyzes the ketonization of 2-hydroxymuconate  
 CC stereoselectively to yield 2-oxo-3-hexenedioate.  
 CC -!- PATHWAY: 2-hydroxymuconic semialdehyde meta-cleavage pathway.  
 CC -!- PATHWAY: Phenol degradation.  
 CC -!- SUBUNIT: Homohexamer (By similarity).  
 CC -!- SIMILARITY: Belongs to the tautomerase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB029044; BAA88507.1; -.  
 DR HSSP; Q01468; 1BJP.  
 DR InterPro; IPR004370; Taut.  
 DR Pfam; PF01361; Tautomerase; 1.  
 DR ProDom; PD404143; Taut; 1.  
 DR TIGRFAMs; TIGR00013; taut; 1.  
 KW Isomerase; Aromatic hydrocarbons catabolism.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT ACT\_SITE 1 1 CATALYTIC BASE (BY SIMILARITY).  
 SQ SEQUENCE 62 AA; 6831 MW; 92CBDDFDFA734D7 CRC64;

Query Match 17.5%; Score 37; DB 1; Length 62;  
 Best Local Similarity 58.8%; Pred. No. 3.4e+02;  
 Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 18 QKSRVIENPTEALSVAV 34  
 || ||| | || ||  
 Db 15 QKKAVIEKVTRALVEAV 31

#### RESULT 20

YH13\_ARCFU

ID YH13\_ARCFU STANDARD; PRT; 59 AA.  
 AC O28560;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein AF1713.  
 GN AF1713.  
 OS Archaeoglobus fulgidus.



OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
 OC Archaeoglobaceae; Archaeoglobus.  
 OX NCBI\_TaxID=2234;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=98049343; PubMed=9389475;  
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,  
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
 RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon *Archaeoglobus fulgidus*.";  
 RL Nature 390:364-370(1997).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE000985; AAB89543.1; -.  
 DR PIR; H69463; H69463.  
 DR TIGR; AF1713; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 59 AA; 6867 MW; C62D3A1D9DDDFE35 CRC64;

Query Match 17.0%; Score 36; DB 1; Length 59;  
 Best Local Similarity 31.0%; Pred. No. 4.4e+02;  
 Matches 9; Conservative 4; Mismatches 12; Indels 4; Gaps 1;

Qy 18 QKSRVIENPTEALSVAVEE----GLAWRK 42  
 | : : | : | || | : | |  
 Db 24 QEEEISEEEAKELDRIVEETKKNIGIPWEK 52

#### RESULT 21

4OT3\_PSEPU

ID 4OT3\_PSEPU STANDARD; PRT; 62 AA.  
 AC Q9Z431;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE 4-oxalocrotonate tautomerase (EC 5.3.2.-) (4-OT).  
 GN NAHJ.  
 OS Pseudomonas putida.  
 OG Plasmid NAH7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=G7 / ATCC 17485;  
 RX MEDLINE=99255564; PubMed=10322041;  
 RA Grimm A.C., Harwood C.S.;  
 RT "NahY, a catabolic plasmid-encoded receptor required for chemotaxis of  
 RT Pseudomonas putida to the aromatic hydrocarbon naphthalene.";  
 RL J. Bacteriol. 181:3310-3316(1999).  
 CC -!- FUNCTION: Catalyzes the ketonization of 2-hydroxymuconate  
 CC stereoselectively to yield 2-oxo-3-hexenedioate.  
 CC -!- PATHWAY: Salicylate meta-cleavage pathway.  
 CC -!- PATHWAY: Naphthalene degradation.  
 CC -!- SUBUNIT: Homohexamer (By similarity).  
 CC -!- SIMILARITY: Belongs to the tautomerase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF100302; AAD13221.1; -.  
 DR HSSP; P49172; 1OTF.  
 DR InterPro; IPR004370; Taut.  
 DR Pfam; PF01361; Tautomerase; 1.  
 DR ProDom; PD404143; Taut; 1.  
 DR TIGRFAMs; TIGR00013; taut; 1.  
 KW Isomerase; Plasmid; Aromatic hydrocarbons catabolism.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT ACT\_SITE 1 1 CATALYTIC BASE (BY SIMILARITY).  
 SQ SEQUENCE 62 AA; 6991 MW; 2E8FFCBBA328FE62 CRC64;

Query Match 17.0%; Score 36; DB 1; Length 62;  
 Best Local Similarity 26.5%; Pred. No. 4.6e+02;  
 Matches 9; Conservative 7; Mismatches 18; Indels 0; Gaps 0;

Qy 3 RSISENSLVAMDFSGQKSRVIENPTEALSVAVEE 36  
 || : : : | || :: | | : | : |  
 Db 11 RSDEQKETLIREVSEAMSRSLDAPIERVVRVIITE 44

#### RESULT 22

##### 4OT\_PSEFL

ID 4OT\_PSEFL STANDARD; PRT; 62 AA.  
 AC Q8KRR5;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE 4-oxalocrotonate tautomerase (EC 5.3.2.-) (4-OT).  
 GN NAHJ.  
 OS Pseudomonas fluorescens.  
 OG Plasmid pLP6a.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=294;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LP6a;  
 RA McFarlane D.M., Foght J.M.;  
 RT "Nucleotide sequence from the lower pathway of naphthalene degradation  
 RT in Pseudomonas fluorescens LP6a.";  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Catalyzes the ketonization of 2-hydroxymuconate  
 CC stereoselectively to yield 2-oxo-3-hexenedioate.  
 CC -!- PATHWAY: Salicylate meta-cleavage pathway.  
 CC -!- PATHWAY: Naphthalene degradation.  
 CC -!- SUBUNIT: Homoheptamer (By similarity).  
 CC -!- SIMILARITY: Belongs to the tautomerase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF525494; AAM88237.1; -.  
 DR InterPro; IPR004370; Taut.  
 DR Pfam; PF01361; Tautomerase; 1.  
 DR ProDom; PD404143; Taut; 1.  
 DR TIGRFAMs; TIGR00013; taut; 1.  
 KW Isomerase; Plasmid; Aromatic hydrocarbons catabolism.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT ACT\_SITE 1 1 CATALYTIC BASE (BY SIMILARITY).  
 SQ SEQUENCE 62 AA; 6976 MW; 7F8347D6184938B9 CRC64;

Query Match 17.0%; Score 36; DB 1; Length 62;  
 Best Local Similarity 26.5%; Pred. No. 4.6e+02;  
 Matches 9; Conservative 7; Mismatches 18; Indels 0; Gaps 0;

Qy 3 RSISENSLVAMDFSGQKSRVIENTEALSVAVEE 36  
 || : : : | || :: | | : | : |  
 Db 11 RSNEQKETLIREVSEAMSRSLDAPIERVVITE 44

#### RESULT 23

##### CSP7\_STRCL

ID CSP7\_STRCL STANDARD; PRT; 66 AA.  
 AC Q01761;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Cold shock-like protein 7.0.  
 GN SC7.0.  
 OS Streptomyces clavuligerus.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1901;  
 RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-40.  
 RC STRAIN=ATCC 27064 / DSM 738 / NRRL 3585;  
 RX MEDLINE=93065223; PubMed=1437568;  
 RA Av-Gay Y., Aharonowitz Y., Cohen G.;  
 RT "Streptomyces contain a 7.0 kDa cold shock like protein.";  
 RL Nucleic Acids Res. 20:5478-5478(1992).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- INDUCTION: In response to low temperature.  
 CC -!- SIMILARITY: Belongs to the cold-shock domain (CSD) family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X68245; CAA48316.1; -.  
 DR PIR; S26378; S26378.  
 DR HSSP; P41016; 1C90.  
 DR InterPro; IPR002059; Cold\_shock.  
 DR InterPro; IPR008994; Nucleic\_acid\_OB.  
 DR Pfam; PF00313; CSD; 1.  
 DR PRINTS; PR00050; COLDSHOCK.  
 DR ProDom; PD000621; Cold\_shock; 1.  
 DR SMART; SM00357; CSP; 1.  
 DR PROSITE; PS00352; COLD\_SHOCK; 1.  
 KW Transcription regulation; DNA-binding; Activator.  
 FT DOMAIN 4 63 CSD.  
 SQ SEQUENCE 66 AA; 7016 MW; CCD5C7858FEB4707 CRC64;

Query Match 17.0%; Score 36; DB 1; Length 66;  
 Best Local Similarity 33.3%; Pred. No. 5e+02;  
 Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

Qy 3 RSISENSLVAMDFSGQKSRVIENPTEA 29  
 ||: ||: | | : : ||: |  
 Db 40 RSLEENQVVNFVDVTHGEGPQAENVSPA 66

#### RESULT 24

##### IPKG\_HUMAN

ID IPKG\_HUMAN STANDARD; PRT; 76 AA.  
 AC Q9Y2B9;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE cAMP-dependent protein kinase inhibitor, gamma form (PKI-gamma).  
 GN PKIG.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Saito T., Miyajima N.;

RL Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Parathyroid;  
 RX MEDLINE=97364742; PubMed=9218452;  
 RA Collins S.P., Uhler M.D.;  
 RT "Characterization of PKI-gamma, a novel isoform of the protein kinase  
 RT inhibitor of cAMP-dependent protein kinase.";  
 RL J. Biol. Chem. 272:18169-18178(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21060778; PubMed=10880337;  
 RA Zheng L., Yu L., Tu Q., Zhang M., He H., Chen W., Gao J., Yu J.,  
 RA Wu Q., Zhao S.;  
 RT "Cloning and mapping of human PKIB and PKIG, and comparison of tissue  
 RT expression patterns of three members of the protein kinase inhibitor  
 RT family, including PKIA.";  
 RL Biochem. J. 349:403-407(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21638749; PubMed=11780052;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Lehvaeslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,  
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 CC -!- FUNCTION: Extremely potent competitive inhibitor of cAMP-dependent  
 CC protein kinase activity, this protein interacts with the catalytic  
 CC subunit of the enzyme after the cAMP-induced dissociation of its  
 CC regulatory chains (By similarity).  
 CC -!- TISSUE SPECIFICITY: Ubiquitous.  
 CC -!- SIMILARITY: Belongs to the PKI family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AB019517; BAA77336.1; -.  
DR EMBL; AF182032; AAD55445.1; -.  
DR EMBL; Z97053; CAC18874.1; -.  
DR Genew; HGNC:9019; PKIG.  
DR MIM; 604932; -.  
DR GO; GO:0004862; F:cAMP-dependent protein kinase inhibitor act. . .; TAS.  
DR InterPro; IPR004171; cAMP\_dep\_PKI.  
DR Pfam; PF02827; PKI; 1.  
DR ProDom; PD010366; cAMP\_dep\_PKI; 1.  
KW Protein kinase inhibitor.  
SQ SEQUENCE 76 AA; 7910 MW; F01B4C73ED2CC6EE CRC64;

Query Match 17.0%; Score 36; DB 1; Length 76;  
Best Local Similarity 31.0%; Pred. No. 5.8e+02;  
Matches 9; Conservative 10; Mismatches 8; Indels 2; Gaps 1;

Qy 6 SENSLVAMDFSGQKSRV--IENPTEALSV 32  
| : :: | :|::: | | : :||:|  
Db 7 SYSDFISCDRTGRRNAVPDIQGDSEAVSV 35

#### RESULT 25

##### IPKG\_MOUSE

ID IPKG\_MOUSE STANDARD; PRT; 76 AA.  
AC O70139;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE cAMP-dependent protein kinase inhibitor, gamma form (PKI-gamma).  
GN PKIG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J;  
RX MEDLINE=97364742; PubMed=9218452;  
RA Collins S.P., Uhler M.D.;  
RT "Characterization of PKI-gamma, a novel isoform of the protein kinase  
RT inhibitor of cAMP-dependent protein kinase.";  
RL J. Biol. Chem. 272:18169-18178(1997).  
CC -!- FUNCTION: Extremely potent competitive inhibitor of cAMP-dependent  
CC protein kinase activity, this protein interacts with the catalytic  
CC subunit of the enzyme after the cAMP-induced dissociation of its  
CC regulatory chains (By similarity).  
CC -!- SIMILARITY: Belongs to the PKI family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U97170; AAC09065.1; -.  
DR MGD; MGI:1343086; Pkig.  
DR InterPro; IPR004171; cAMP\_dep\_PKI.  
DR Pfam; PF02827; PKI; 1.  
DR ProDom; PD010366; cAMP\_dep\_PKI; 1.  
KW Protein kinase inhibitor.  
SQ SEQUENCE 76 AA; 7943 MW; 965F577D80C8DE59 CRC64;

Query Match 17.0%; Score 36; DB 1; Length 76;  
Best Local Similarity 31.0%; Pred. No. 5.8e+02;  
Matches 9; Conservative 10; Mismatches 8; Indels 2; Gaps 1;

Qy 6 SENSLVAMDFSGQKSRV--IENPTEALSV 32  
| : :: | :|::: | |: :||:|  
Db 7 SYSDFISCDRTGRRNAVDPDIQGDSEAVSV 35

Search completed: July 8, 2004, 08:20:05  
Job time : 16.5433 secs